YAYASAN PRIMA AGUS TEKNIK

ARTIFICIAL INTELLIGENCE DAN DATA MINING

Dalam Kerangka Sekuriti

Dr. Joseph Teguh Santoso, S.Kom, M.Kom.

ARTIFICIAL INTELLIGENCE DAN DATA MINING

Dalam Kerangka Sekuriti

Dr. Joseph Teguh Santoso, S.Kom, M.Kom.

YAYASAN PRIMA AGUS TEKNIK

PENERBIT :
YAYASAN PRIMA AGUS TEKNIK
Jl. Majapahit No. 605 Semarang
Telp. (024) 6723456. Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

ISBN 978-634-7227-26-3 (PDF)

ARTIFICIAL INTELLIGENCE DAN DATA MINING
Dalam Kerangka Sekuriti

Penulis :
Dr. Joseph Teguh Santoso, S.Kom., M.Kom.

ISBN : 978-634-7227-26-3

Editor :
Dr. Agus Wibowo, M.Kom, M.Si, MM.

Penyunting :
Dr. Mars Caroline Wibowo. S.T., M.Mm.Tech

Desain Sampul dan Tata Letak :
Irdha Yunianto, S.Ds., M.Kom

Penebit :
Yayasan Prima Agus Teknik Bekerja sama dengan
Universitas Sains & Teknologi Komputer (Universitas STEKOM)

Anggota IKAPI No: 279 / ALB / JTE / 2023

Redaksi :
Jl. Majapahit no 605 Semarang
Telp. 08122925000
Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

Distributor Tunggal :
Universitas STEKOM
Jl. Majapahit no 605 Semarang
Telp. 08122925000
Fax. 024-6710144
Email : info@stekom.ac.id

KATA PENGANTAR

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa, buku **"Artificial Intelligence Dan Data Mining dalam Kerangka Sekuriti"** ini dapat diterbitkan sebagai respons terhadap kebutuhan akan literatur yang komprehensif di bidang keamanan siber berbasis kecerdasan buatan. Di tengah pesatnya revolusi digital dan bertambah rumitnya ancaman dunia maya, penggunaan teknologi cerdas seperti AI tidak lagi bersifat opsional, melainkan menjadi pilar utama pertahanan sistem digital modern.

Buku ini hadir sebagai kumpulan wawasan, metode, dan studi kasus yang membahas peran AI secara mendalam dalam memberikan solusi atas tantangan keamanan siber masa kini dan masa depan. Di dalamnya, pembaca akan menemukan pemaparan mulai dari konsep dasar, arsitektur sistem, hingga aplikasi nyata AI dalam menjaga privasi data, mendeteksi intrusi, memitigasi serangan, serta mengoptimalkan mekanisme monitoring berbasis IoT. Tidak hanya itu, isu-isu mutakhir seperti keamanan multi-tenancy, integrasi fuzzy systems, dan pendekatan data mining juga diuraikan secara aplikatif guna memperkaya pemahaman dan kapasitas implementasi.

Pada setiap bab, penulis menghadirkan pembahasan yang sistematis agar pembaca—baik mahasiswa, dosen, peneliti, maupun praktisi—dapat memperoleh pencerahan tentang bagaimana AI mengubah paradigm keamanan digital. Rangkaian contoh nyata seperti deteksi botnet, filtering spam, pengenalan wajah dengan ILPB-SVM, hingga inspeksi perangkat medis berbasis IoT dan jaringan saraf tiruan semakin menegaskan peran vital AI dalam beragam aspek keamanan siber.

Penulis berharap, pemaparan dalam buku ini mampu menjadi jembatan pengetahuan antara teori dan praktik, sekaligus membekali pembaca dengan sudut pandang kritis dan solutif dalam pengembangan teknologi keamanan siber yang adaptif, inovatif, dan kompetitif di era AI.

Ucapan terima kasih penulis haturkan kepada para rekan akademisi, praktisi, serta semua pihak yang telah memberikan dukungan materiil maupun moril hingga terwujudnya buku ini. Semoga kehadiran buku ini menjadi sumber inspirasi dan referensi tepercaya dalam meningkatkan kapabilitas keamanan digital bangsa.

Selamat membaca dan semoga bermanfaat!

Semarang, Juli 2025
Penulis

Dr. Joseph Teguh Santoso, S.Kom., M.Kom.

DAFTAR ISI

BAB 1
PERAN AI DALAM KEAMANAN SIBER

1.1 PENDAHULUAN

Kecerdasan Buatan (AI) dapat dikarakterisasikan sebagai pengambilan keputusan buatan yang mirip dengan pengambilan keputusan manusia, berdasarkan algoritma unik tertentu dan estimasi matematika terkait. Keamanan Siber berkaitan dengan langkah-langkah yang diambil untuk melindungi dari serangan digital di dunia virtual. Selain itu, pekerjaan AI terus berkembang di dunia modern, di mana terdapat ancaman yang membayangi keamanan siber. Dengan kemajuan dalam inovasi, kejahatan siber juga meningkat dan menjadi tidak dapat diprediksi. Penjahat siber meluncurkan serangan canggih yang membahayakan kerangka kerja keamanan saat ini. Dengan demikian, bisnis keamanan siber berkembang untuk memenuhi kebutuhan keamanan organisasi yang terus berkembang. Namun, strategi pertahanan profesional keamanan ini mungkin tidak memenuhi harapan dan mungkin gagal memenuhi agenda yang diusulkan cepat atau lambat.

Kebutuhan akan Kecerdasan Buatan

Pekerjaan vital AI adalah mengalihkan pekerjaan dari insinyur keamanan siber manusia saat ini, untuk menangani kedalaman dan detail yang tidak dapat ditangani manusia secara efektif. Kemajuan dalam teknologi pembelajaran mesin menyiratkan bahwa aplikasi AI juga dapat secara otomatis beradaptasi dengan perubahan ancaman dan menemukan masalah saat muncul. Kebutuhan keamanan siber yang dapat dipenuhi oleh perangkat dan platform AI:

- **Luas Data:**
  Orang-orang langsung bingung saat dihadapkan dengan sejumlah besar informasi log dan peringatan yang diberikan oleh kerangka kerja saat ini. Pemrograman kecerdasan buatan yang berjalan pada prosesor canggih saat ini dapat memproses lebih banyak data dalam hitungan menit daripada yang dapat ditangani manusia dalam hitungan bulan. Dengan demikian, ia juga dapat memperhitungkan masalah dan ketidakkonsistenan sambil menangani sejumlah besar informasi keamanan.
- **Jarum ancaman:**
  Perburuan ancaman siber adalah pencarian proaktif yang konstan melalui jaringan dan kumpulan data untuk mendeteksi ancaman yang menghindari perangkat komputerisasi yang ada.
  Pelanggar hukum digital kini berada di dalam berbagai kerangka kerja, menunggu untuk menyelesaikan serangan mereka. Mereka sering kali dapat melarikan diri dari orang-orang. Namun, AI dapat dengan cepat memeriksa berbagai keadaan untuk mendeteksi jarum ancaman dibandingkan dengan aktivitas jahat.
- **Pengoptimalan Respons:**
  Kecerdasan buatan dapat mempercepat pengenalan masalah yang dapat dibuktikan, dengan cepat merujuk silang berbagai peringatan dan sumber informasi keamanan.

Prioritas insiden yang akan ditangani akan tetap menjadi domain pakar keamanan siber manusia, tetapi mereka dapat dibantu lebih lanjut oleh sistem AI yang akan meningkatkan kecepatan pengenalan dan waktu reaksi.

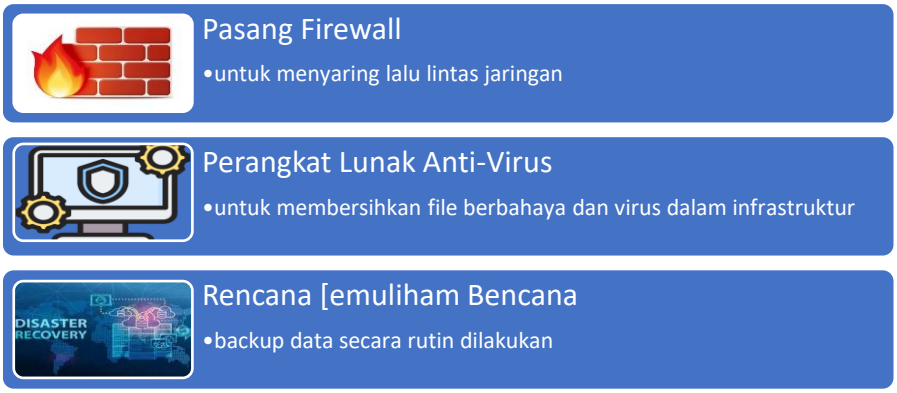
- **Perlombaan senjata AI:**
  Penjahat dunia maya saat ini sudah dilengkapi dengan teknik AI yang canggih. Teknologi AI secara umum dapat menjadi berkah atau kutukan. Programmer dapat dengan mudah memanfaatkan alat-alat terbaru untuk meluncurkan serangan yang lebih canggih, yang masing-masing lebih berbahaya. Hal ini telah menjadi perlombaan senjata di mana AI menjadi eksponen utama di kedua sisi.

1.2 KECERDASAN BUATAN DALAM KEAMANAN SIBER

AI dalam keamanan siber mendukung perusahaan atau organisasi, memungkinkan mereka untuk menjaga mekanisme pertahanan mereka; lebih jauh lagi, AI membantu mereka untuk menafsirkan kejahatan siber secara efektif. Perusahaan menggunakan peluang ideal ini untuk mencapai efisiensi dalam otomatisasi dengan beralih ke digital karena mereka memanfaatkan kecepatan eksekusi yang lebih cepat. Mencapai konektivitas digital dalam seluruh rantai nilai mereka membantu mereka untuk memenuhi persaingan yang semakin ketat di pasar. Pada jalur yang sama, penjahat siber menemukan peluang dengan meningkatnya digitalisasi. Serikat penjahat siber secara aktif berfokus pada ekosistem digital termasuk infrastruktur cloud, perangkat *Internet of Things* (IoT), dan penawaran perangkat lunak sebagai layanan / *Software as a Service* (SaaS). Oleh karena itu, Perusahaan dihadapkan dengan tantangan untuk mendorong keuntungan yang lebih besar dalam keuntungan bisnis sambil menyeimbangkan risiko paparan siber.

Desain Sistem Keamanan Berlapis-lapis

Organisasi lebih berkonsentrasi pada keamanan siber dalam skenario saat ini. Hal ini dikarenakan serangan keamanan siber yang canggih telah memaksa mereka untuk mengeluarkan banyak uang guna mencegah terjadinya pelanggaran data di masa mendatang. Dimulai dengan merancang kerangka kerja keamanan berlapis yang akan mengamankan infrastruktur jaringan.



Gambar 1.1 Infrastruktur Jaringan

Gambar 1.1 menunjukkan infrastruktur jaringan yang berisi Firewall, perangkat lunak anti-virus, dan rencana pemulihan bencana. Semua komponen ini membuat infrastruktur jaringan lebih efisien. AI telah memengaruhi keamanan dengan membantu para ahli mengenali ketidaknormalan dalam sistem dengan menganalisis aktivitas klien dan merenungkan contoh-contohnya. Para ahli keamanan kini dapat merenungkan dan mengatur informasi dengan memanfaatkan AI dan mendeteksi kerentanan untuk mencegah serangan berbahaya.

Pendekatan Keamanan Tradisional dan AI

AI akan membantu meningkatkan pendekatan keamanan tradisional dengan cara-cara berikut:

- Instrumen keamanan canggih bertenaga AI akan digunakan untuk menyaring dan bereaksi terhadap peristiwa keamanan.
- Firewall modern akan memiliki teknologi pembelajaran mesin bawaan untuk mendeteksi dan menghapus pola yang tidak biasa dalam lalu lintas sistem, jika dianggap berbahaya.
- Dengan menganalisis kerentanan menggunakan fitur pemrosesan bahasa alami dalam AI, para ahli keamanan juga dapat mengidentifikasi akar serangan digital.
- Diperlukan analisis prediktif untuk mendeteksi ancaman berbahaya dan pemindaian data terlebih dahulu.

Karena ketergantungan kita pada big data meningkat, kita telah menciptakan kebutuhan paralel untuk menjaganya tetap aman. Dengan demikian, kebutuhan saat ini adalah untuk menjaga integritas jaringan, data yang tersimpan, dan program dari akses dan serangan yang tidak sah

1.3 PEKERJAAN TERKAIT

Internet digunakan oleh jutaan orang biasa, sehingga mereka menjadi sasaran empuk bagi penjahat dunia maya. Dengan "perangkat lunak dan digitalisasi" serta adopsi IoT yang cepat, keamanan dunia maya kini menjadi inti dari strategi bisnis. Data merupakan kategorisasi yang luas, mulai dari informasi kartu kredit, bank, catatan keuangan, dan informasi pribadi. Solusi kontemporer untuk masalah yang luas ini terletak pada kesadaran dasar, membangun kemampuan dunia maya yang defensif atau perlindungan dan perawatan, melalui pendidikan.

Onashoga, S. Adebukola, Ajayi, O. Bamidele, dan A. Taofik (2013) dalam makalah mereka membahas simulasi arsitektur berbasis multiagen untuk sistem deteksi intrusi guna mengatasi kekurangan sistem deteksi intrusi berbasis agen seluler saat ini. Data didistribusikan pada host dan jaringan. Algoritma penambangan pola tertutup (CPM) diperkenalkan untuk membuat profil aktivitas pengguna dalam basis data jaringan. Hal ini tidak hanya membantu mengurangi waktu penyortiran data tetapi juga membantu analis untuk mengetahui pola perilaku manusia secara real time.

Alex Roney Mathew dkk. (2010) dalam makalah mereka membahas berbagai jenis kejahatan dunia maya, yaitu: phishing rekayasa sosial, spoofing email, dan pharming. Mereka juga membahas cara melindungi orang dari kejahatan tersebut dengan penekanan pada

biometrik. Kejahatan dunia maya akhir-akhir ini menjadi sangat umum sehingga hanya sebagian kecil populasi di dunia yang tidak tersentuh olehnya.

Selvakani, Maheshwari V. dan Karavanisundari (2010) dalam makalah mereka menekankan fakta bahwa teknologi informasi dapat digunakan untuk pekerjaan yang merusak maupun membangun, tergantung pada siapa yang menggunakannya. Studi ini membahas pentingnya hukum dunia maya untuk melindungi kepentingan korban dunia maya. Para penulis percaya bahwa komputer dapat diamankan bahkan oleh orang dengan pengetahuan sederhana, tetapi bahwa penetapan dan pelestarian bukti merupakan tugas yang sulit. Diperlukan hukum yang harmonis antara teknologi dan hukum; diperlukan kombinasi yang baik. AI harus membantu dalam merancang hukum yang kuat yang dapat digunakan secara efektif untuk melacak kejahatan dunia maya.

L.S. Wijesinghe, L.N.B. De Silva, G.T.A. Abhayaratne, P. Krithika, S.M.D.R. Priyashan, dan Dhishan Dhammearatchi (2016) dalam makalah penelitian mereka terutama berfokus pada cara memerangi kejahatan dunia maya, dan juga menjelaskan seberapa cerdas dan efektif alat "agen" dapat digunakan dalam pendeteksian dan pencegahan serangan dunia maya. Serangan dunia maya cenderung berdampak besar pada industri TI dalam hal pencurian data, data tersebut.

Ramamoorthy R. (2010) dalam makalahnya membahas berbagai perspektif keamanan siber. Karena ancaman baru yang terus berkembang terhadap perusahaan, TI telah menjadikan keamanan siber sebagai masalah yang "harus diperhatikan". Tim administrasi sistem harus merancang cara untuk meningkatkan keamanan siber mereka dengan solusi pengujian keamanan aplikasi otomatis, sesuai permintaan, dan waktu nyata yang membuat keamanan siber komprehensif untuk aplikasi menjadi lebih sederhana dan lebih hemat biaya. Keamanan siber tidak mengenal batas. Penulis menyinggung tentang pengendalian penyebaran server untuk meningkatkan efisiensi operasional dan memudahkan pemulihan bencana, virtualisasi jelas memberikan hasil akhir.

Yasmin N., dan Bajaj N. (2012) dalam makalah penelitian mereka menyajikan Modifikasi S-box dalam DES. DES adalah Standar Enkripsi Data dan "kotak pengganti" S-box perangkat enkripsi standar. Keamanan merupakan perhatian utama bagi organisasi yang berpartisipasi dalam pertukaran informasi. Salah satu aspek penting untuk komunikasi yang aman adalah kriptografi. Karena kejahatan dunia maya menyebabkan kerugian finansial yang serius, sistem yang ada memerlukan modifikasi terus-menerus untuk memastikan bahwa tingkat keamanan tidak terganggu. Hal ini menunjukkan tingkat ketahanan yang lebih tinggi terhadap serangan pada hubungan $Li+1 = Ri$. Namun, diperlukan sejumlah besar pengetahuan matematika dan pemahaman tentang sistem kriptografi yang lengkap.

Dampak

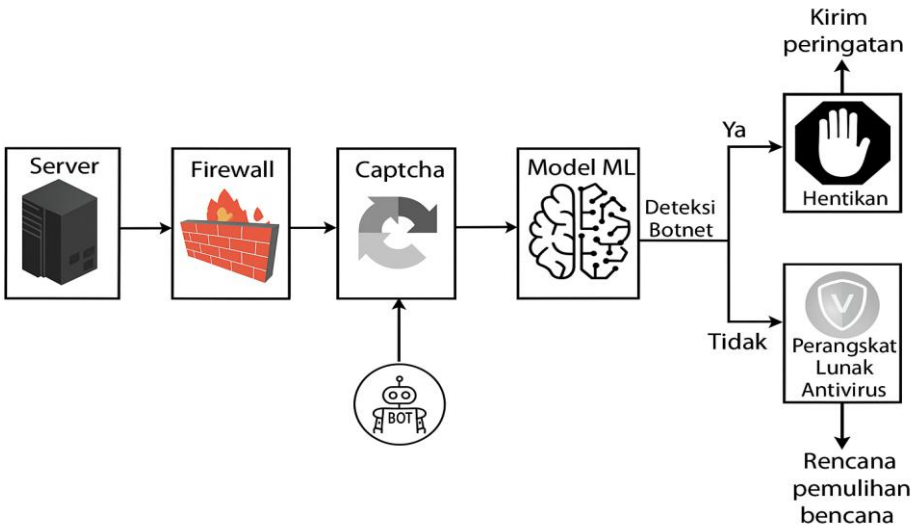
Serangan siber meluas dengan cepat, meskipun langkah-langkah keamanan telah ditingkatkan. Serangan tersebut dapat berupa malware, serangan phishing, pencurian kata sandi, serangan Trojan, dan sebagainya. Untuk menghindari kejahatan siber ini, langkah-langkah keamanan siber yang kuat diperlukan. Teknologi yang sedang berkembang seperti ilmu kognitif, komputasi awan, robotika, perbankan internet, dan e-commerce sangat perlu

menerapkan langkah-langkah perlindungan yang memadai dalam domain keamanan siber. Dengan meningkatnya penggunaan Big Data, yang mengatur pengambilan keputusan dengan menggunakan model pembelajaran mesin, keamanan siber perlu menjadi yang terdepan

Usulan Pekerjaan

Model ini mengusulkan untuk menambahkan lapisan keamanan ke pendekatan keamanan berlapis. Arsitektur sistem yang diusulkan dijelaskan dalam Gambar 1.2:

1. Misalkan, saat kita mencoba masuk ke rekening bank menggunakan kredensial kita, sebuah bot mencoba memecahkan captcha.
2. Setiap kali hal itu terjadi, model pembelajaran mesin yang didasarkan pada kemampuannya mengenali pola dari masa lalu akan mendeteksi keberadaan bot melalui pemantauan aktif dan analisis prediktif.
3. Jika terdeteksi, model tersebut akan menghentikan proses saat ini dan mengirimkan peringatan.
4. Jika bot tidak ada, maka model tersebut akan melanjutkan proses dan menjalankan perangkat lunak anti-virus, untuk menghapus file berbahaya lainnya.
5. Rencana pemulihan bencana pada akhirnya akan memastikan bahwa data penting apa pun tidak hilang dan dicadangkan.



Gambar 1.2 Arsitektur Sistem

1.4 ARSITEKTUR SISTEM

Cakupan Masa Depan

Meskipun kita merangkul cara-cara baru interaksi digital dan lebih banyak infrastruktur penting kita yang beralih ke digital, parameter transformasi yang sedang berlangsung belum dipahami oleh sebagian besar dari kita. Pemahaman yang lebih baik tentang arsitektur dunia maya global diperlukan.

AI menemukan penerapannya di hampir setiap bidang sains dan teknik. Model AI memerlukan perlindungan yang tepat dalam keamanan digital dan teknologi baru untuk

melawan pembelajaran mesin yang antagonis, mempertahankan kerahasiaan, dan mengamankan pembelajaran yang terorganisasi, dan sebagainya. Dalam bab ini, penulis meneliti pendekatan khusus dalam AI yang menjanjikan dan mengusulkan sistem untuk mencegah jenis serangan keamanan siber tertentu.

BAB 2
MELINDUNGI PRIVASI DATA

Di era digital saat ini, teknik penambangan data (*data mining*) memiliki peran penting dalam mengekstraksi pengetahuan tersembunyi dari kumpulan data berskala besar. Teknik ini membantu mengidentifikasi pola, tren, dan hubungan yang sebelumnya tidak terlihat, yang sangat berguna dalam pengambilan keputusan. Namun, di balik manfaat tersebut, terdapat potensi ancaman terhadap privasi individu, terutama ketika data pribadi dianalisis tanpa persetujuan atau pengamanan yang memadai.

Bab ini bertujuan untuk membahas secara komprehensif berbagai teknik penambangan data, aplikasinya dalam meningkatkan keamanan sistem informasi, serta potensi risiko yang menyertainya. Salah satu teknik penting yang dibahas adalah klasifikasi, yaitu proses mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang terdeteksi.

Dalam konteks keamanan siber, teknik klasifikasi sangat berguna untuk:

- Mengidentifikasi dan memisahkan pengguna normal dari pengguna jahat berdasarkan aktivitas mereka di media sosial atau sistem informasi.
- Membedakan antara situs web resmi dan situs phishing dengan menganalisis karakteristik konten dan struktur domain.
- Memberikan peringatan dini kepada pengguna ketika mereka hendak menjalankan skrip atau kode yang terindikasi berbahaya, dengan cara melabeli kode tersebut sebagai ancaman.

Selain itu, salah satu aplikasi paling krusial dari data mining adalah deteksi intrusi. Melalui penerapan algoritma seperti *decision tree, support vector machine* (SVM), atau deep learning, sistem dapat memantau lalu lintas jaringan dan mendeteksi aktivitas mencurigakan secara real-time. Ketika potensi serangan teridentifikasi, sistem dapat secara otomatis memberikan laporan dan mengambil langkah mitigasi, seperti pemblokiran IP atau penonaktifan akses pengguna.

Untuk menjaga keseimbangan antara pemanfaatan data mining dan perlindungan privasi, penting untuk menerapkan prinsip-prinsip seperti *privacy-preserving data mining* (PPDM), di mana data dapat ditambang tanpa mengekspos informasi sensitif pengguna.

Dengan demikian, meskipun teknik penambangan data membawa banyak manfaat dalam bidang keamanan, penting untuk selalu mengimbanginya dengan kebijakan privasi, regulasi yang tepat, dan pendekatan teknologi yang etis.

.

2.1 PENDAHULUAN

Sistem komputer memiliki kemampuan untuk melindungi informasi berharga, data mentah beserta sumber dayanya dalam hal privasi, kebenaran, dan keaslian; kemampuan ini dikenal sebagai keamanan komputer. Pihak ketiga tidak dapat membaca atau mengedit konten basis data dengan menggunakan parameter, yaitu Privasi/kerahasiaan dan integritas. Dengan

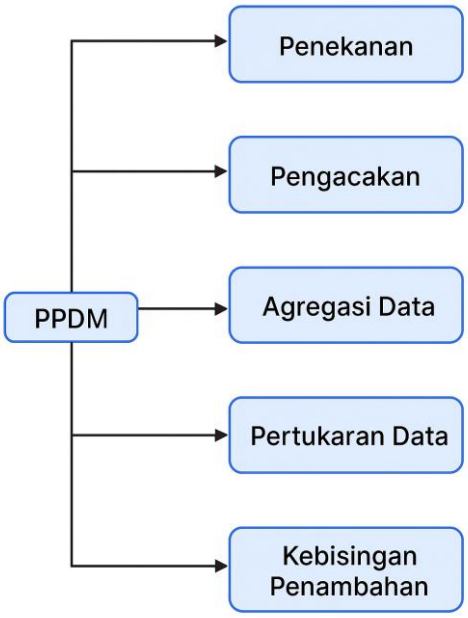
menggunakan parameter keaslian, orang yang tidak berwenang tidak diperbolehkan untuk mengubah, menggunakan, atau melihat konten basis data. Ketika satu atau lebih sumber daya komputer membahayakan ketersediaan, integritas, atau kerahasiaan melalui suatu tindakan, hal itu dikenal sebagai intrusi.

Jenis serangan ini dapat dicegah dengan menggunakan firewall dan kebijakan penyaringan router. Intrusi dapat terjadi bahkan dalam sistem yang paling aman dan oleh karena itu disarankan untuk mendeteksinya sejak awal. Dengan menggunakan teknik penambangan data, pola fitur suatu sistem dapat dideteksi oleh sistem deteksi intrusi (IDS) sehingga anomali dapat dideteksi dengan bantuan serangkaian pengklasifikasi yang sesuai. Untuk mendeteksi intrusi dengan mudah, beberapa teknik penambangan data penting seperti klasifikasi dan pengelompokan sangat membantu.

Data uji dapat dianalisis dan diberi label ke dalam jenis kelas yang diketahui dengan bantuan teknik klasifikasi. Untuk pengelompokan objek ke dalam serangkaian kluster, metode pengelompokan digunakan. Metode ini digunakan sedemikian rupa sehingga kluster memiliki semua objek yang serupa. Mungkin ada beberapa tantangan keamanan untuk penambangan pengetahuan yang mendasarinya dari sejumlah besar data serta ekstraksi pola tersembunyi dengan menggunakan teknik penambangan data. Untuk mengatasi masalah ini, *Privacy Preserving Data Mining* (PPDM) digunakan, yang bertujuan untuk memperoleh informasi penting dan berguna dari basis data yang tidak diinginkan atau informal.

a) Penekanan

Informasi pribadi atau sensitif seseorang seperti nama, gaji, alamat, dan usia, jika ditekan sebelum perhitungan apa pun dikenal sebagai penekanan. Penekanan dapat dilakukan dengan bantuan beberapa teknik seperti Pembulatan (Rs/- 15365.87 dapat dibulatkan menjadi 15.000), Bentuk lengkap (Nama Chitra Mehra dapat diganti dengan inisial, yaitu, CM dan Tempat India dapat diganti dengan IND dan seterusnya). Ketika ada persyaratan akses penuh ke nilai-nilai sensitif, penekanan tidak dapat digunakan oleh penambangan data. Cara lain untuk melakukan penekanan adalah dengan membatasi daripada menekan informasi sensitif catatan. Metode yang dapat kita gunakan untuk menekan hubungan identitas suatu catatan disebut sebagai De-identifikasi. Salah satu teknik de-identifikasi tersebut adalah k-Anonymity. Jaminan perlindungan data yang dirilis terhadap identifikasi ulang atau de-identifikasi seseorang. K-anonimitas dan penerapannya sulit dilakukan sebelum mengumpulkan data lengkap di satu tempat tepercaya. Untuk solusinya, solusi kriptografi berbasis teknik berbagi rahasia dapat digunakan.



Gambar 2.1 Pendekatan Penambangan Data Yang Menjaga Privasi.

b) Pengacakan Data

Server pusat suatu organisasi mengambil informasi dari banyak pelanggan dan membangun model agregat dengan melakukan berbagai teknik penambangan data. Hal ini memungkinkan pelanggan untuk menyajikan gangguan yang tepat atau mengganggu catatan secara acak dan untuk mengetahui informasi yang akurat dari kumpulan data tersebut. Ada beberapa cara untuk memasukkan gangguan, yaitu, penambahan atau perkalian nilai yang dihasilkan secara acak. Untuk mencapai pelestarian privasi yang diperlukan, kami menggunakan agitasi dalam teknik pengacakan data. Untuk menghasilkan catatan individual, gangguan yang dihasilkan secara acak dapat ditambahkan ke data inovatif. Gangguan yang ditambahkan ke data asli tidak dapat dipulihkan dan dengan demikian mengarah pada privasi yang diinginkan.

Berikut ini adalah langkah-langkah teknik pengacakan:

1. Setelah data diacak oleh penyedia data, data tersebut akan disampaikan ke Penerima Data.
2. Dengan menggunakan algoritma rekonstruksi distribusi, penerima data dapat melakukan perhitungan distribusi pada data yang sama.

c) Agregasi Data

Data digabungkan dari berbagai sumber untuk memfasilitasi analisis data dengan teknik agregasi data. Dengan melakukan ini, penyerang dapat menyimpulkan data tingkat pribadi dan individu dan juga mengenali sumber daya tersebut. Ketika data yang diekstraksi memungkinkan penambang data untuk mengidentifikasi individu tertentu, privasi penambang data dianggap berada dalam ancaman serius. Ketika data dianonimkan segera setelah proses agregasi, data tersebut dapat dicegah untuk diidentifikasi, meskipun, kumpulan data yang

dianonimkan tersebut memuat informasi yang cukup yang diperlukan untuk identifikasi individu.

d) Pertukaran Data

Demi perlindungan privasi, pertukaran nilai di berbagai catatan dapat dilakukan dengan menggunakan proses ini. Privasi data masih dapat dipertahankan dengan memungkinkan perhitungan agregat dicapai persis seperti yang dilakukan sebelumnya, yaitu, tanpa mengganggu total orde yang lebih rendah dari data. K-anonimitas dapat digunakan dalam kombinasi dengan teknik ini serta dengan kerangka lain untuk melanggar definisi privasi model tersebut.

e) Penambahan/Gangguan Noise

Untuk akurasi kueri yang maksimal dan mengurangi peluang identifikasi rekamannya, terdapat mekanisme yang disediakan dengan penambahan noise yang terkontrol. Berikut ini adalah beberapa teknik yang digunakan untuk penambahan noise:

1. Komposisi Paralel
2. Mekanisme Laplace
3. Komposisi Berurutan

2.2 TEKNIK PENAMBANGAN DATA DAN PERANNYA DALAM KLASIFIKASI DAN DETEKSI

Program komputer malware yang berulang kali menyebar dari satu komputer ke komputer lain disebut worm. Malware terdiri dari *adware, worm, Trojan horse, virus komputer, spyware, key logger, http worm, UDP worm dan port scan worm*, dan *remote to local worm*, kode berbahaya lainnya dan *user to root worm*. Ada berbagai alasan mengapa penyerang menulis program ini, seperti:

i. Proses komputer dan interupsinya
ii. Perakitan informasi sensitif
iii. Sistem pribadi dapat memperoleh akses

Sangat penting untuk mendeteksi worm di internet karena dua alasan berikut:

- Worm menciptakan titik-titik rentan
- Kinerja sistem dapat menurun

Oleh karena itu, penting untuk memperhatikan worm sejak awal dan mengkategorikannya dengan bantuan algoritma klasifikasi penambangan data. Berikut ini adalah algoritma klasifikasi yang dapat digunakan; Jaringan Bayesian, Random Forest, Decision Tree, dll. Prinsip yang mendasarinya adalah bahwa sistem deteksi intrusi (IDS) dapat digunakan oleh sebagian besar teknik deteksi worm. Sangat sulit untuk memprediksi bentuk worm selanjutnya. Ada tantangan dalam deteksi otomatis worm dalam sistem. Sistem Deteksi Intrusi secara umum dapat diklasifikasikan menjadi dua jenis:

i. Berdasarkan jaringan: paket jaringan dipantulkan hingga saat itu kecuali jika tidak disebarkan ke host akhir
ii. Berdasarkan host: Paket jaringan yang telah disebarkan ke host akhir dipantulkan

Lebih jauh, paket jaringan yang dikodekan merupakan area inti dari deteksi berbasis host IDS untuk mendeteksi serangan worm internet. Kita harus memperhatikan kinerja lalu lintas dalam

jaringan dengan berfokus pada paket jaringan tanpa pengkodean. Banyak teknik pembelajaran mesin telah digunakan untuk sistem deteksi worm dan intrusi. Dengan demikian, teknik penambangan data dan pembelajaran mesin sangat penting karena keduanya memiliki peran penting dalam sistem deteksi worm.

Banyak model Deteksi Intrusi telah diusulkan dengan menggunakan berbagai skema penambangan data. Untuk mempelajari garis besar yang tidak teratur dan tidak biasa dari set pelatihan, Pohon Keputusan dan Algoritma Genetika Pembelajaran Mesin dapat digunakan dan kemudian berdasarkan pengklasifikasi yang dihasilkan, dapat diberi label sebagai kelas Normal atau Abnormal untuk data uji. Data berlabel, "Abnormal", berguna untuk menunjukkan keberadaan intrusi.

a) Pohon Keputusan

Salah satu teknik pembelajaran mesin yang paling populer adalah teknik pohon keputusan Quinlan. Sejumlah keputusan dan simpul daun diperlukan untuk membangun pohon dengan mengikuti teknik bagi-dan-taklukkan. Suatu kondisi perlu diuji dengan menggunakan atribut data masukan dengan bantuan setiap simpul keputusan untuk menangani hasil pengujian yang terpisah. Dalam pohon keputusan, kita memiliki sejumlah cabang. Simpul daun direpresentasikan oleh hasil keputusan. Set data pelatihan T memiliki satu set n-kelas {C1, C2,..., Cn} ketika set data pelatihan T terdiri dari kasus-kasus yang termasuk dalam satu kelas, maka ia diperlakukan sebagai daun. T juga dapat diperlakukan sebagai daun jika T kosong tanpa kasus. Jumlah hasil pengujian dapat dilambangkan dengan k jika ada k subset dari T yaitu {T1, T2, ..., Tk}, di mana. Proses ini berulang pada setiap Tj, di mana 1 <= j <= n, hingga setiap subset tidak termasuk dalam satu kelas. Saat membangun pohon keputusan, pilih atribut terbaik untuk setiap simpul keputusan.

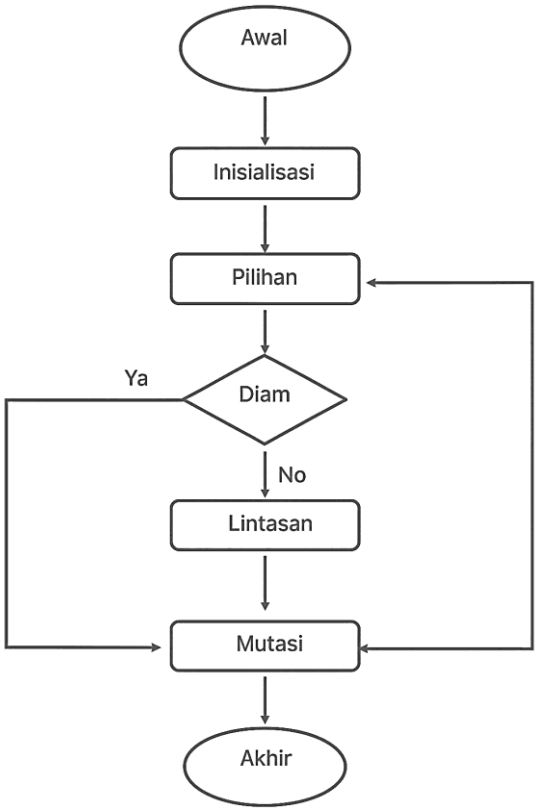
Kriteria Rasio Keuntungan diadopsi oleh Pohon Keputusan C4.5. Dengan menggunakan kriteria ini, atribut yang memberikan keuntungan informasi maksimum dengan mengurangi uji bias/favoritisme dipilih. Jadi, untuk mengklasifikasikan data uji, pohon yang dibangun digunakan yang fitur dan fitur data pelatihannya sama. Persetujuan pengujian di atas dapat dilakukan dengan memulai dari simpul akar. Atas dasar hasil, cabang yang mengarah ke anak harus diikuti. Proses akan diulang secara rekursif untuk waktu hingga anak bukan daun. Untuk memeriksa kelas dan daun yang sesuai, kasus uji harus diterapkan.

b) Algoritma Genetika (GA)

Digunakan untuk memecahkan masalah dengan menggunakan teknik evolusi biologis dengan bantuan pendekatan pembelajaran mesin. Populasi solusi kandidat dapat dioptimalkan dengan bantuan Algoritma Genetika. Dalam algoritma genetika, operator genetika, yaitu seleksi, persilangan, dan mutasi membantu pemodelan struktur data pada kromosom. Pada awalnya, pembangkitan acak populasi kromosom dapat dilakukan. Dengan cara ini, akan ada semua kemungkinan solusi dari suatu masalah dalam populasi dan itu dianggap sebagai solusi kandidat.

Lokasi kromosom yang berbeda disebut "gen" yang dapat ditentukan sebagai angka, karakter, atau bit. Untuk mengevaluasi kebaikan setiap kromosom berdasarkan solusi yang diinginkan, kami menggunakan fungsi kebugaran. Reproduksi alami dapat dirangsang oleh

operator persilangan sedangkan mutasi spesies dirangsang oleh operator mutasi. Kromosom yang paling cocok dapat dipilih oleh operator seleksi. Algoritma Genetika dan operasinya dapat direpresentasikan pada Gambar 2.2. Berikut ini adalah tiga faktor penting yang harus kita pertimbangkan sebelum menggunakan algoritma genetika untuk memecahkan berbagai masalah.



Gambar 2.2 Diagram Alir Algoritma Genetik.

1. Fungsi kebugaran
2. Representasi individu
3. Parameter algoritma genetik

Untuk merancang sistem kekebalan buatan, metode berbasis algoritma genetik dapat digunakan. Dengan menggunakan metode ini, metode untuk mendeteksi malware pada ponsel pintar telah diusulkan oleh Bin et al. Dalam pendekatan ini, tanda tangan statis dan dinamis dari malware diekstraksi untuk mendapatkan skor berbahaya dari sampel yang diuji.

c) Random Forest

Ini adalah algoritma klasifikasi yang menggunakan kumpulan pengklasifikasi terstruktur pohon. Dalam algoritma ini, kelas dipilih sebagai kelas pemenang berdasarkan suara yang diberikan oleh pohon individu dari hutan. Untuk membangun pohon, ada persyaratan data acak dari kumpulan data pelatihan. Dengan demikian, kumpulan data yang dipilih dapat dibagi menjadi kumpulan data pelatihan dan kumpulan data pengujian. Data pelatihan mencakup sebagian besar kumpulan data sedangkan data pengujian akan memiliki sebagian kecil dari kumpulan data. Berikut ini adalah langkah-langkah yang diperlukan untuk konstruksi pohon:

1. Sampel kasus N dipilih secara acak dari kumpulan data asli yang mewakili set pelatihan yang diperlukan untuk menumbuhkan pohon.
2. Dari M variabel input, m variabel dapat dipilih secara acak. Nilai m akan konstan pada saat menumbuhkan hutan.
3. Nilai maksimum yang mungkin dapat diberikan untuk setiap pohon di hutan. Tidak ada persyaratan pemangkasan atau Pemangkasan pohon.
4. Untuk membentuk hutan acak, semua pohon klasifikasi dapat digabungkan. Masalah overfitting pada kumpulan data besar dapat diperbaiki dengan bantuan hutan acak. Hutan acak juga dapat melatih/ menguji dengan cepat pada kumpulan data yang kompleks. Hutan acak juga dapat disebut sebagai teknik penambangan Data Operasional.

Setiap pohon klasifikasi dapat digunakan untuk memberikan suara untuk suatu kelas karena fitur khususnya. Berdasarkan suara maksimum yang diberikan ke suatu kelas, kelas solusi dibangun.

d) Penambangan aturan asosiasi

Digunakan untuk menemukan hubungan yang menarik di antara sekumpulan atribut dalam kumpulan data. Aturan asosiasi dapat didefinisikan sebagai hubungan antar kumpulan data. Aturan asosiasi sangat membantu untuk membangun keputusan strategis tentang berbagai tindakan seperti manajemen rak, penetapan harga promosi, dan masih banyak lagi. Sebelumnya, seorang analis data terlibat dalam penambangan aturan asosiasi yang tugasnya adalah menemukan pola atau aturan asosiasi dalam kumpulan data yang diberikan kepadanya. Analisis yang canggih pada kumpulan data yang sangat besar ini dapat dicapai dengan cara yang hemat biaya, tetapi mungkin ada risiko keamanan data) bagi pemilik data karena penambang data dapat menambang informasi yang sensitif. Saat ini, dalam penemuan data pengetahuan (KDD), penambangan aturan asosiasi digunakan secara luas untuk penemuan pola. Masalah (ARM) dapat dipecahkan dengan menavigasi item dalam database dengan bantuan berbagai algoritma berdasarkan kebutuhan pengguna. Algoritma penambangan aturan asosiasi (ARM) dapat diklasifikasikan secara luas menjadi DFS (*Depth First Search*) dan BFS (*Breadth First Search*) berdasarkan pendekatan yang digunakan untuk melintasi ruang pencarian.

Kedua metode ini, yaitu DFS (*Depth First Search*) dan BFS (*Breadth First Search*) selanjutnya dibagi menjadi metode - *intersecting* dan *counting*, berdasarkan set item dan nilai dukungannya. Algoritma Apriori-DIC, Apriori dan Apriori-TID adalah algoritma strategi penghitungan berbasis BFS, sedangkan algoritma partisi adalah algoritma BFS strategi intersecting. *Algoritma Equivalence Class Clustering and bottom-up Lattice Traversal* (ECLAT) bekerja pada strategi interseksional dengan DFS. DFS dengan strategi Counting terdiri dari algoritma FP-Growth,. Untuk peningkatan kecepatan, algoritma ini dapat dioptimalkan secara khusus,. *Breadth First Search* (BFS) dengan Menghitung Kejadian: Algoritma yang menonjol dalam kelompok ini adalah algoritma Apriori. Dengan memotong kandidat dengan subset yang langka dan dengan bantuan algoritma ini, properti penutupan ke bawah dari suatu itemset dapat dimanfaatkan.

Hal ini harus dilakukan sebelum menghitung dukungannya. Dua parameter penting yang harus diukur pada saat evaluasi aturan asosiasi yaitu: dukungan dan keyakinan. Dalam BFS, dimungkinkan untuk melakukan optimasi yang diinginkan dengan mengetahui nilai dukungan dari semua subset kandidat terlebih dahulu. Kelemahan utama dari yang disebutkan di atas adalah peningkatan kompleksitas komputasi dalam aturan yang telah diekstraksi dari basis data yang besar. Bentuk algoritma Apriori yang lebih baik, tersebar, dan tidak aman adalah algoritma *Fast Distributed Mining* (FDM). Organisasi dapat menggunakan data dengan lebih kompeten dengan bantuan kemajuan dalam teknik penambangan data.

Dimungkinkan dalam Apriori untuk menghitung kandidat dengan kardinalitas k dengan bantuan pemindaian tunggal dari basis data yang besar. Keterbatasan terpenting dari algoritma apriori adalah mencari kandidat dalam setiap transaksi. Untuk melakukan hal yang sama, struktur pohon hash digunakan. Ekstensi Apriori, yaitu Apriori-TID, menandakan kandidat terkini yang menjadi dasar setiap transaksi, sementara basis data mentah sudah cukup untuk Apriori normal. Apriori dan Apriori-TID bila digabungkan membentuk Apriori-Hybrid. Pohon awalan digunakan untuk memperbaiki pemisahan yang terjadi antara proses, penghitungan, dan pembuatan kandidat dalam Apriori-DIC.

2.3 PENGELOMPOKAN

Teknik penambangan data digunakan untuk mengelompokkan sekumpulan objek sedemikian rupa sehingga terdapat lebih banyak kesamaan pada objek-objek dari kelas yang sama dibandingkan dengan objek-objek dari kelas lainnya. Artinya klaster tersebut berkelas sama, yakni kesamaan intra-klaster bernilai maksimum dan kesamaan inter-klaster bernilai minimum. Pembelajaran tanpa pengawasan dapat dilakukan dengan bantuan pengelompokan. Berikut ini adalah jenis-jenis algoritma pengelompokan :

a) Berbasis Distribusi
b) Berbasis Kepadatan
c) Berbasis Sentroid
d) Berbasis Koneksi atau Pengelompokan Hirarkis
e) Teknik Pengelompokan Terkini

a) Pengelompokan Berbasis Distribusi

Model pengelompokan di mana tanggal dikelompokkan/dipasang dalam model berdasarkan probabilitas, yaitu, dengan cara apa tanggal tersebut dapat dipasang dalam distribusi yang sama. Dengan demikian, kelompok yang dibentuk akan berdasarkan distribusi normal atau distribusi Gaussian

b) Pengelompokan Berbasis Kepadatan

Dalam jenis pengelompokan ini, sebuah klaster dibentuk dengan bantuan area dengan kepadatan lebih tinggi dibandingkan dengan data lainnya. Berikut ini adalah tiga teknik Pengelompokan Berbasis Kepadatan yang paling sering digunakan:

i) Pergeseran Rata-rata
ii) OPTIK
iii) DBSCAN

c) Pengelompokan Berbasis Sentroid

Kluster yang direpresentasikan oleh vektor merupakan bagian dari pengelompokan berbasis sentroid. Bukan merupakan persyaratan wajib bahwa klaster ini harus menjadi bagian dari kumpulan data yang diberikan. Jumlah klaster tidak memadai untuk ukuran k dalam algoritma pengelompokan rata-rata k; oleh karena itu, penting untuk menemukan pusat klaster k dan mengalokasikan objek ke pusat terdekatnya. Dengan mengambil nilai yang berbeda dari k inisialisasi acak, algoritma ini berjalan beberapa kali untuk memilih yang terbaik dari beberapa kali jalan. Dalam pengelompokan medoid k, klaster dibatasi secara ketat pada anggota kumpulan data, sedangkan dalam pengelompokan median k, median diambil untuk membentuk klaster; kelemahan utama dari teknik ini adalah kita harus memilih jumlah klaster terlebih dahulu.

d) Pengelompokan Berbasis Koneksi (Hierarkis)

Sesuai dengan namanya, pengelompokan jenis ini dilakukan berdasarkan kedekatan atau jarak objek. Poin kunci terpenting untuk membentuk jenis pengelompokan ini adalah jarak antara objek yang dapat dihubungkan satu sama lain dan membentuk pengelompokan. Alih-alih membagi dataset secara tunggal, algoritme ini menyediakan hierarki mendalam untuk menggabungkan pengelompokan pada jarak tertentu. Untuk merepresentasikan pengelompokan, digunakan dendrogram. Jarak penggabungan pengelompokan ditunjukkan pada sumbu y dan penempatan objek menunjukkan sumbu x untuk memastikan bahwa pengelompokan tidak tercampur. Berdasarkan cara berbeda dalam menghitung jarak, ada beberapa jenis klaster berbasis koneks i:

i) Klaster Tautan Tunggal
ii) Tautan Lengkap
iii) Klaster Tautan Rata-rata

e) Teknik Klaster Terbaru

Untuk data berdimensi tinggi, teknik pengelompokan standar yang disebutkan di atas tidak sesuai, oleh karena itu beberapa teknik baru sedang ditemukan. Teknik baru ini dapat diklasifikasikan ke dalam dua kategori utama, yaitu: Pengelompokan Subruang dan Pengelompokan Korelasi. Daftar kecil atribut yang harus diukur untuk pembentukan klaster dipertimbangkan dalam pengelompokan subruang. Korelasi antara atribut yang dipilih juga dapat dilakukan dengan pengelompokan korelasi.

2.4 PENAMBANGAN DATA YANG MENJAGA PRIVASI (PPDM)

Untuk mengekstraksi pengetahuan yang relevan dari sejumlah besar data dan untuk melindungi semua informasi sensitif dari basis data tersebut, kami menggunakan penambangan data yang menjaga privasi (PPDM). Teknik-teknik ini dibuat dengan tujuan untuk memastikan perlindungan data sensitif sehingga privasi dapat dilindungi dengan kinerja yang efisien dari semua operasi penambangan data. Ada dua kelas teknik penambangan data yang berkaitan dengan privasi:

1. Privasi data
2. Privasi informasi

Modifikasi basis data untuk perlindungan data sensitif individu, kami menggunakan teknik privasi data. Jika ada persyaratan untuk modifikasi pengetahuan sensitif yang dapat disimpulkan dari basis data, teknik privasi informasi lebih disukai. Untuk memberikan privasi pada input, privasi data lebih disukai, sedangkan untuk memberikan privasi pada output, teknik privasi informasi digunakan. Untuk melindungi informasi pribadi dari paparan adalah fokus utama dari algoritma PPDM.

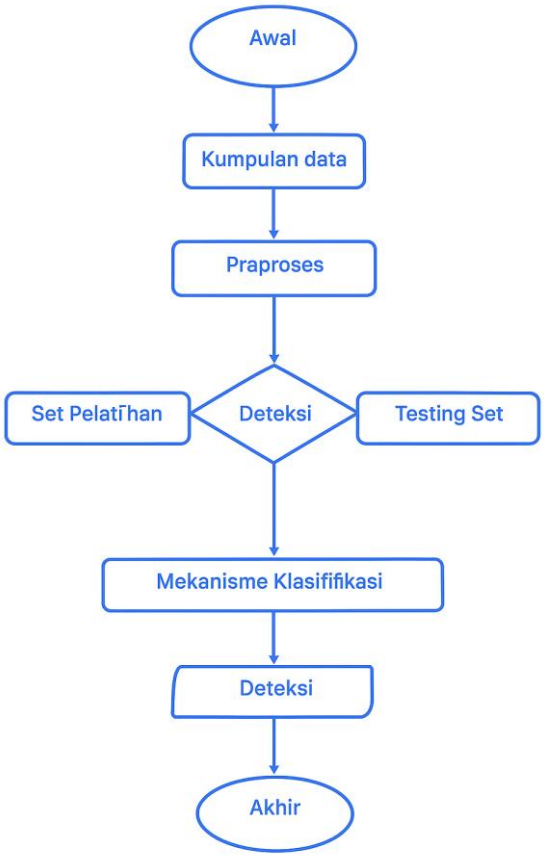
Hal ini bergantung pada analisis algoritma penambangan yang dicapai selama privasi data. Tujuan utama Penambangan Data Pelestarian Privasi adalah membangun algoritma yang mengubah data asli dalam beberapa cara yang berguna, sehingga tidak ada visibilitas data dan pengetahuan pribadi bahkan setelah proses penambangan berhasil. Undang-undang privasi akan mengizinkan akses jika beberapa manfaat terkait yang memuaskan ditemukan sebagai hasil dari akses tersebut.

2.5 SISTEM DETEKSI INTRUSI (IDS)

Deteksi awal intrusi adalah tujuan utama dari sistem deteksi intrusi. Ada persyaratan pengetahuan manusia tingkat tinggi dan sejumlah besar waktu untuk mencapai keamanan dalam penambangan data. Namun, sistem deteksi intrusi berdasarkan penambangan data membutuhkan lebih sedikit keahlian untuk kinerja yang lebih baik. Untuk memahami serangan jaringan yang berbeda dengan layanan yang rentan, sistem deteksi intrusi sangat membantu. Serangan berbasis data berbasis aplikasi selalu meningkatkan hak istimewa, login yang tidak sah dan aksesibilitas file sangat sensitif.

Proses penambangan data dapat digunakan sebagai alat untuk keamanan siber untuk deteksi malware yang kompeten dari kode. Gambar 2.3 menunjukkan garis besar sistem deteksi intrusi. Beberapa komponen seperti sensor, monitor konsol, dan mesin pusat membentuk sistem deteksi intrusi yang lengkap. Peristiwa keamanan dihasilkan oleh sensor sedangkan tugas monitor konsol adalah memantau dan mengendalikan semua peristiwa dan peringatan. Fungsi utama mesin pusat adalah merekam peristiwa dalam basis data dan berdasarkan peristiwa ini, peringatan dapat dibuat diikuti oleh serangkaian aturan tertentu. Faktor-faktor berikut bertanggung jawab atas klasifikasi sistem deteksi intrusi:

- ✓ Lokasi
- ✓ Jenis Sensor
- ✓ Teknik yang digunakan oleh mesin pusat untuk menghasilkan peringatan.



Gambar 2.3 Tinjauan Umum Sistem Deteksi Intrusi (Ids).

Ketiga komponen sistem deteksi intrusi dapat diintegrasikan ke dalam satu perangkat.

Jenis-jenis IDS

Deteksi intrusi dapat dilakukan baik pada jaringan maupun dengan sistem individual dan oleh karena itu terdapat tiga jenis IDS, yaitu: Berbasis Jaringan, Berbasis Host, dan IDS Hibrida.

IDS Berbasis Jaringan

Jaringan komputer telah menjadi target musuh dan penjahat karena perannya yang semakin dinamis dalam masyarakat modern. Sangat penting untuk menemukan solusi terbaik demi melindungi sistem kita. Berbagai teknik pencegahan intrusi seperti menghindari kesalahan pemrograman, perlindungan informasi menggunakan teknik enkripsi dan biometrik atau kata sandi dapat digunakan sebagai garis depan keamanan. Dengan menggunakan teknik pencegahan intrusi sebagai satu-satunya tindakan perlindungan, sistem kita tidak 100% aman dari serangan tempur. Untuk memberikan keamanan tambahan bagi sistem komputer, teknik-teknik yang disebutkan di atas digunakan.

Berbagai sumber daya seperti akun pengguna, sistem berkas mereka, dan kernel sistem dari sistem target harus dilindungi oleh sistem deteksi intrusi. Untuk sistem deteksi intrusi berbasis jaringan, sumber data adalah paket-paket jaringan. Untuk mendengarkan dan menganalisis lalu lintas jaringan saat paket-paket berjalan melintasi jaringan, sistem deteksi

intrusi berbasis jaringan (NIDS) menggunakan adaptor jaringan. Sistem deteksi intrusi berbasis jaringan digunakan untuk menghasilkan peringatan untuk mendeteksi intrusi yang berada di luar batas perusahaan.

Berikut ini adalah keuntungan dari IDS Berbasis Jaringan:

1. IDS dapat dibuat tidak terlihat untuk meningkatkan keamanan terhadap serangan.
2. Jaringan berukuran besar dapat dipantau oleh IDS berbasis jaringan.
3. IDS ini dapat memberikan hasil yang lebih baik tanpa mengganggu kerja jaringan yang biasa.
4. IDS mudah dipasang ke dalam jaringan yang sudah ada.

Keterbatasan IDS Berbasis Jaringan adalah sebagai berikut:

a) Informasi terenkripsi jaringan privat virtual tidak dapat dianalisis dengan IDS berbasis jaringan.
b) Implementasi IDS berbasis jaringan yang berhasil didasarkan pada sakelar perantara yang ada di jaringan.
c) IDS berbasis jaringan tidak akan stabil dan akan mogok ketika penyerang memecah paket mereka dan melepaskannya.

IDS Berbasis Host

Pada jenis IDS ini, berbagai log dapat disaring dengan bantuan sensor yang ditempatkan pada sumber daya jaringan. Log ini dibuat oleh sistem operasi host atau program aplikasi. Peristiwa atau tindakan tertentu yang mungkin terjadi pada sumber daya jaringan individual dicatat oleh log audit. Jenis IDS ini dapat menangani bahkan serangan yang tidak dapat ditangani. Karena itu, penyerang dapat menyalahgunakan salah satu orang dalam yang tepercaya.

Basis aturan tanda tangan yang berasal dari kebijakan keamanan yang khusus untuk suatu situs digunakan oleh sistem berbasis host. Semua masalah yang terkait dengan IDS Berbasis Jaringan dapat diatasi oleh IDS berbasis host karena dapat memberi tahu personel keamanan dengan detail lokasi penyusupan. Dengan demikian, orang tersebut dapat mengambil tindakan langsung untuk menghentikan penyusupan.

Berikut ini adalah keuntungan dari IDS Berbasis Host:

- Dapat mendeteksi bahkan serangan yang tidak terdeteksi oleh IDS Berbasis Jaringan.
- Untuk mendeteksi serangan yang menyangkut pelanggaran integritas perangkat lunak, ia bekerja pada jejak log audit sistem operasi.

Kekurangan Host-Based IDS adalah sebagai berikut:

- Berbagai jenis serangan DoS (*Denial of Service*) dapat menonaktifkan Host-Based ID.
- Serangan yang menargetkan jaringan tidak dapat dideteksi oleh host-based IDS.
- Mengonfigurasi dan mengelola setiap sistem individual sangatlah sulit.

Hybrid IDS

Ini adalah kombinasi dari jaringan dan host-based IDS untuk membentuk struktur bagi sistem deteksi intrusi generasi berikutnya. Pengaturan ini secara umum dikenal sebagai sistem deteksi intrusi fusi/hibrida. Dengan menambahkan jaringan dan host-based IDS, ia akan secara signifikan meningkatkan ketahanan terhadap beberapa serangan lagi. Teknik penambangan

data yang diperlukan untuk IDS adalah Pencocokan Pola, Klasifikasi dan Pemilihan Fitur Pencocokan Pola.

2.6 KLASIFIKASI SITUS WEB PHISHING

Ini adalah jenis serangan rekayasa sosial yang umumnya digunakan untuk mencuri data pengguna, seperti kredensial login dan nomor kartu kredit. Untuk menutupi situs web yang jujur, situs web palsu biasanya dibuat oleh orang-orang yang melakukan penipuan. Karena aktivitas phishing dari penyerang, pengguna secara keliru kehilangan uang mereka. Oleh karena itu, langkah penting harus diambil untuk melindungi perdagangan daring. Kebaikan fitur yang diekstraksi menunjukkan keakuratan prediksi dan klasifikasi situs web. Alat anti-phishing digunakan oleh sebagian besar pengguna internet untuk merasa aman terhadap serangan phishing. Alat anti-phishing diperlukan untuk memprediksi phishing yang akurat. Bagian konten situs web phishing bersama dengan indikator keamanan mungkin memiliki serangkaian petunjuk di dalam browser.

Berbagai metode telah diusulkan untuk menangani masalah phishing. Untuk memprediksi serangan phishing, klasifikasi berbasis aturan, yang merupakan teknik penambangan data, digunakan sebagai metode yang efektif untuk prediksi. Jika penyerang mengirim email kepada korban dengan meminta mereka untuk mengungkapkan informasi pribadi mereka, itu merupakan indikasi phishing. Untuk membuat situs web phishing dengan trik yang tepat, serangkaian fitur bersama digunakan oleh para phishing. Kita dapat membedakan antara situs web phishing dan non-phishing berdasarkan fitur yang diekstraksi dari situs web yang dikunjungi tersebut.

Identifikasi situs phishing dapat dilakukan dengan bantuan dua pendekatan :

i) Berbasis daftar hitam: Ini mencakup analisis komparatif URL, yaitu, yang diminta bersama dengan URL lain yang ada dalam daftar tersebut.

ii) Berbasis heuristik: Fitur-fitur tertentu dari berbagai situs web dikumpulkan dan diberi label sebagai phishing atau asli.

Kelemahan utama dari pendekatan daftar hitam adalah tidak dapat mencakup semua situs web phishing karena setiap detik, situs web jahat baru diluncurkan, sementara pendekatan berbasis heuristik dapat mengidentifikasi situs web palsu yang asli. Metode berbasis heuristik bergantung pada pemilihan fitur dan cara pemrosesannya. Penambangan data digunakan untuk menemukan hubungan dan pola di antara fitur-fitur dalam kumpulan data tertentu. Tugas utama penambangan data adalah mengambil keputusan karena keputusan-keputusan ini bergantung pada pola dan aturan yang telah diturunkan menggunakan algoritma penambangan data. Meskipun kemajuan yang cukup besar telah dibuat untuk pengembangan teknik pencegahan, phishing masih merupakan ancaman karena teknik yang digunakan untuk penanggulangan masih didasarkan pada daftar hitam URL reaktif. Karena masa pakai situs web phishing yang lebih pendek, metode yang digunakan di situs-situs ini dianggap tidak efektif. Pendekatan baru, klasifikasi asosiatif (AC) ditemukan lebih tepat untuk jenis aplikasi ini; pendekatan ini merupakan campuran dari aturan Asosiasi dan teknik Klasifikasi penambangan data.

Ada dua tahap dalam klasifikasi asosiasi (AC):

i) Fase pelatihan: Tahap ini digunakan untuk menginduksi pengetahuan tersembunyi (aturan) dengan bantuan aturan Asosiasi.

ii) Fase klasifikasi: Tahap ini digunakan untuk membangun pengklasifikasi setelah memotong aturan yang tidak efektif dan berlebihan.

Telah dibuktikan dari banyak studi penelitian bahwa pengklasifikasi asosiasi (AC) secara umum menunjukkan pengklasifikasi yang lebih baik dalam hal tingkat kesalahan daripada pohon keputusan dan induksi aturan (pendekatan klasifikasi standar).

2.7 SERANGAN DENGAN MITIGASI INJEKSI KODE

Serangan injeksi kode adalah teknik untuk menulis kode mesin baru ke dalam memori program yang rentan. Jika ada bug dalam program, kontrol dapat dikirim ke kode baru setelah memanipulasinya. $W + X$, teknik perlindungan meringankan serangan injeksi kode dengan mengizinkan satu operasi, yaitu, untuk menulis atau mengeksekusi operasi tetapi tidak keduanya secara bersamaan.

Penyuntikan Kode dan Kategorinya

Berikut ini adalah jenis-jenis serangan penyuntikan kode:

i) **Penyuntikan SQL**: Dapat didefinisikan sebagai teknik yang menggunakan sintaks SQL untuk memasukkan perintah guna membaca, mengubah, atau memodifikasi basis data. Misalnya, ada kolom pada halaman web mengenai autentikasi kata sandi pengguna. Umumnya, kita menggunakan kode skrip untuk ini. Kode skrip ini akan menghasilkan kueri SQL sehingga kata sandi yang dimasukkan sesuai dengan daftar nama pengguna dapat diverifikasi:

ii) **HTML Script Injection**: Kode berbahaya dapat disuntikkan oleh penyerang dengan bantuan tag. Dengan demikian, properti lokasi dokumen akan diubah dengan menyetelnya ke skrip yang disuntikkan.

iii) **Object Injection**: Hypertext pre-processor (PHP) digunakan untuk serialisasi dan deserialisasi objek. Dengan bantuan injeksi objek, kelas yang ada dalam program dapat dimodifikasi dan serangan berbahaya dapat dijalankan jika input yang tidak dapat dipercaya diizinkan masuk ke fungsi deserialisasi.

iv) **Remote File Injection**: Untuk menyebabkan kerusakan yang diinginkan, nama file yang terinfeksi dari jarak jauh dapat diberikan oleh penyerang dengan mengubah perintah jalur file skrip sebagai jalur.

v) **Code Reuse Attacks**: Serangan penggunaan ulang kode (CRA) merupakan perkembangan terbaru dalam keamanan. Serangan ini terjadi ketika penyerang mengekspresikan aliran kontrol melalui kode yang sudah ada sebelumnya. Dengan menggunakan ini, penyerang diizinkan untuk menjalankan kode acak pada mesin yang disusupi. Ini adalah pendekatan pemrograman berorientasi pengembalian dan berorientasi lompatan. Pendekatan ini dapat mengklaim kembali fragmen kode pustaka. *Return Into Lib C* (RILC) adalah jenis serangan penggunaan kembali kode di mana tumpukan dikompromikan dan kontrol ditransfer ke awal fungsi pustaka yang

ada seperti mprotect() untuk membuat wilayah memori yang memungkinkan operasi penulisan dan eksekusi di atasnya untuk melewati W+X. Untuk mengatasi serangan tersebut, kami menggunakan teknik penambangan data. Ketika kode sumber diperiksa untuk mengungkap kesalahan tersebut dan untuk ini instruksi diklasifikasikan sebagai berbahaya. Beberapa algoritma klasifikasi yang dapat digunakan dalam hal ini adalah Regresi Logistik, Bayesian, Support Vector Machine, dan Pohon Keputusan.

Tujuan utama dari Bab ini adalah untuk menemukan peran teknik Data Mining dalam mencapai keamanan. Beberapa aplikasi seperti *Privacy Preserving Data Mining* (PPDM), *Intrusion Detection System* (IDS), *Phishing Website Classification dan Mitigation of Code Injection* dibahas. Beberapa algoritma *Classification* dan *Clustering* juga dibahas karena perannya yang signifikan dalam sistem deteksi intrusi. Teknik Data Mining dasar lainnya yang digunakan untuk sistem deteksi intrusi seperti *Feature Extraction, Association Rule Mining* dan *Decision Trees* juga dibahas. Aplikasi keamanan lain dari Data Mining seperti *Malware Detection, Spam Detection, Web Mining* dan *Crime Profiling* juga dapat dieksplorasi dalam hal keamanan sebagai cakupan masa depan.

BAB 3
AI UNTUK KEAMANAN SIBER

Kecerdasan Buatan (AI) telah menjadi istilah yang sangat populer di era digital modern. Meski masih dalam tahap pengembangan, AI telah menunjukkan dampak besar terhadap berbagai aspek kehidupan manusia. Saat ini, sulit membayangkan dunia tanpa AI karena teknologi ini telah terintegrasi dalam hampir setiap bidang kehidupan mulai dari game, pengolahan bahasa, pengenalan suara, robotika, hingga transaksi keuangan. Secara umum, tujuan utama AI adalah menciptakan sistem yang mampu meniru cara manusia berpikir, belajar, mengambil keputusan, dan memecahkan masalah. Dengan kata lain, AI berusaha memahami dan mereplikasi pola pikir manusia untuk menyelesaikan tugas-tugas tertentu secara efisien.

Teknologi ini telah menjadi bagian penting dalam mendukung aktivitas manusia, tetapi di sisi lain juga menghadirkan tantangan baru, terutama dalam hal keamanan siber. Serangan digital yang mengincar pemerintah, bank, dan lembaga besar lainnya menunjukkan bahwa ancaman keamanan semakin kompleks. Oleh karena itu, AI juga digunakan untuk memperkuat sistem keamanan digital, seperti mendeteksi anomali dan serangan siber secara otomatis. Namun, pemanfaatan AI dalam bidang keamanan tidak lepas dari kendala. Kompleksitas data, sifat masalah yang dinamis, dan dampaknya terhadap privasi manusia menjadikan pengembangan AI sebagai tantangan tersendiri. Maka dari itu, meskipun AI memiliki potensi besar, perlu ada pendekatan yang bijak dan bertanggung jawab dalam pengembangannya agar tidak menimbulkan risiko baru di masa depan.

3.1 PENDAHULUAN

Kecerdasan Buatan (AI)

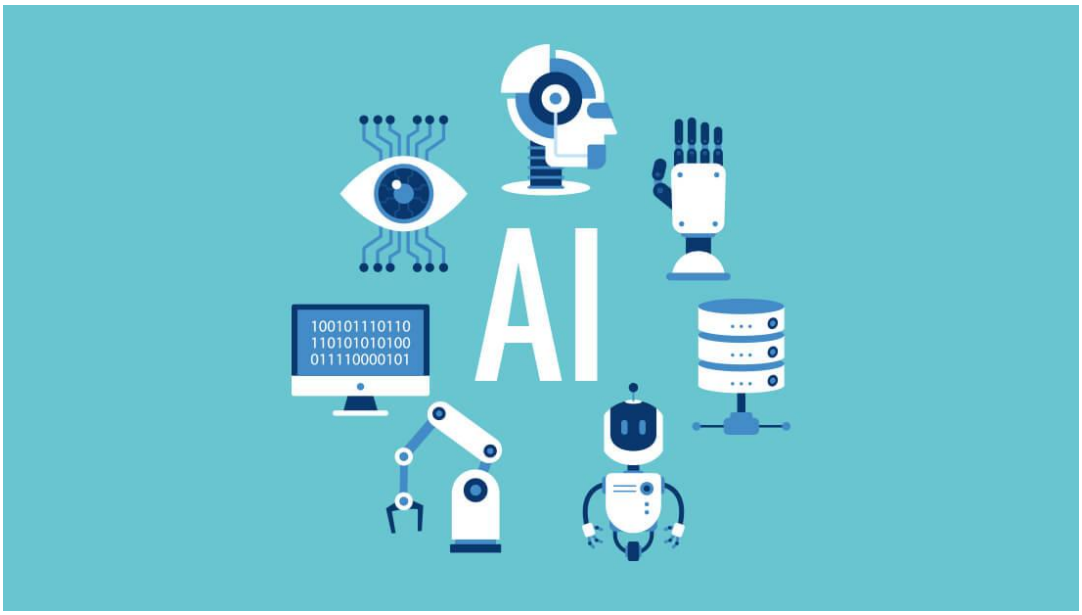
Sistem Kecerdasan Buatan dapat digunakan untuk memahami data yang mengganggu dan melibatkan profesional keamanan untuk memahami kondisi tingkat lanjut sehingga dapat menganalisis tindakan yang tidak teratur. Kecerdasan buatan juga dapat menguntungkan keamanan siber dengan menciptakan strategi robotik setiap kali bahaya digital dikenali. Kekuatan otak AI dapat menganalisis sejumlah besar data dan memungkinkan pengembangan kerangka kerja dan perangkat lunak yang ada dengan cara yang tepat untuk mengurangi serangan digital. Pada dasarnya, penggunaan AI untuk pengaturan keamanan digital akan membantu melindungi asosiasi dari bahaya digital yang ada dan mengidentifikasi jenis malware baru.

Penggabungan kecerdasan buatan ke dalam struktur keamanan dapat digunakan untuk mengurangi risiko keamanan tingkat lanjut yang terus meningkat yang sedang dipertimbangkan oleh asosiasi secara keseluruhan. Melalui usaha tersebut, aplikasi yang menggunakan Pembelajaran Mesin dan juga pemikiran elektronik (AI) secara menyeluruh digunakan secara lebih luas karena pengelompokan data, batas penyimpanan, dan daya

penanganan terus meningkat. Secara terus-menerus, ukuran data yang sangat besar sulit untuk ditangani oleh individu.

Dengan bantuan kecerdasan buatan, sejumlah besar data kemungkinan besar dapat dikurangi dalam hitungan milidetik, sehingga upaya tersebut dapat dengan mudah mengenali dan juga memulihkan diri dari bahaya. Jelas perlindungan terhadap senjata mekanis yang canggih dapat dikembangkan hanya dengan pemrograman yang cerdas, dan kejadian dalam dua tahun terakhir telah menunjukkan data yang berkembang pesat tentang malware dan senjata elektronik. Dengan kemajuan dalam inovasi, kejahatan digital juga berkembang dan menjadi rumit.

Penjahat digital mendorong serangan canggih yang membahayakan kerangka kerja keamanan saat ini. Dengan demikian, bisnis keamanan siber juga maju untuk memenuhi kebutuhan keamanan organisasi yang terus meningkat. Namun, teknik perlindungan defensif ini juga dapat gagal cepat atau lambat. Untuk meningkatkan permainan mereka dan memperbarui komponen identifikasi kelemahan mereka, organisasi memilih Kecerdasan Buatan (AI). Kesadaran AI dalam Keamanan Siber mendukung organisasi untuk mempertahankan sistem perlindungan mereka. Ini juga membantu mereka dalam memecah pelanggaran digital dengan lebih baik.



Gambar 3.1 Kecerdasan Buatan.

Pengetahuan buatan dalam keamanan digital menguntungkan karena meningkatkan cara otoritas keamanan mengeksplorasi, mempelajari, dan menangani kejahatan dunia maya. Pengetahuan buatan meningkatkan inovasi keamanan canggih yang digunakan asosiasi untuk melawan penjahat dunia maya dan membantu memantau asosiasi dan klien. Di sisi lain, AI dapat sangat membutuhkan sumber daya. Pemikiran Kecerdasan Buatan adalah wilayah pengembangan premium dan usaha dalam jaringan keamanan digital

Gambar 3.1 di atas menggambarkan Kecerdasan Buatan untuk Keamanan Siber, yang mencakup semua faktor yang berkaitan dengan dunia maya. Kecerdasan berbasis komputer

mengacu pada kemajuan yang dapat menghargai, menentukan, dan mengeksekusi, bergantung pada data yang diperoleh dan ditentukan. Kecerdasan buatan bekerja dalam tiga cara berbeda:

1. Informasi terbantu, yang tersedia saat ini, meningkatkan pencapaian yang dicapai orang dan organisasi saat ini.
2. Pengetahuan tertambah, yang berkembang saat ini, memungkinkan orang dan organisasi untuk melakukan hal-hal yang sebelumnya tidak dapat mereka lakukan.
3. Wawasan otonom, yang dibuat untuk prospek, menampilkan mesin yang menindaklanjuti sendiri. Contohnya adalah kendaraan yang dapat mengemudi sendiri, ketika digunakan secara luas.

Meskipun masih dalam tahap awal, AI dapat dikatakan memiliki beberapa tingkat wawasan manusia: penyimpanan informasi eksplisit ruang; instrumen untuk mengamankan informasi baru; dan komponen untuk menggunakan informasi tersebut. AI, kerangka kerja induk, sistem saraf, dan pembelajaran mendalam pada umumnya merupakan model atau bagian dari inovasi AI saat ini.

- Pembelajaran mesin menggunakan pendekatan faktual untuk mengimplementasikan struktur PC guna "belajar" (misalnya, meningkatkan eksekusi secara dinamis) dengan memanfaatkan informasi alih-alih disesuaikan secara khusus. AI bekerja paling baik saat difokuskan pada tugas tertentu, bukan tugas penting yang luas.
- Kerangka kerja ahli adalah program yang dimaksudkan untuk menangani masalah di dalam ruang tertentu. Dengan mencerminkan pemikiran spesialis manusia, mereka mengatasi masalah dan membuat keputusan menggunakan pemikiran berbasis standar yang halus melalui kumpulan informasi yang dikurasi dengan hati-hati.
- Sistem saraf menggunakan pandangan dunia pemrograman yang dimotivasi secara organik yang memberdayakan PC untuk mendapatkan keuntungan dari informasi observasional. Dalam sistem saraf, setiap hub memberikan beban pada infonya yang berbicara tentang seberapa benar atau salahnya hal itu dibandingkan dengan aktivitas yang dilakukan. Hasil akhir kemudian dikontrol oleh keseluruhan beban tersebut.
- Pembelajaran Mendalam adalah kumpulan strategi kecerdasan buatan yang lebih luas yang tunduk pada penggambaran data pembelajaran, bukan hitungan tugas-eksplisit. Saat ini, pengenalan gambar melalui pembelajaran mendalam biasanya lebih unggul bagi manusia, dengan berbagai macam kegunaan, misalnya kendaraan tanpa pengemudi, penelitian penelitian, dan penentuan klinis.

3.2 AI UNTUK KEAMANAN SIBER

Keamanan Siber dan AI berubah menjadi instrumen yang luar biasa untuk keamanan siber dan infrastruktur di sekitarnya. AI berfungsi dengan sangat baik ketika produknya disiapkan pada indeks informasi besar dari gadget keamanan siber, organisasi, dan data apa pun yang berharga untuk menyelesaikan apa pun. AI menargetkan identifikasi kekhasan, kehati-hatian, dan ketegasan. Setiap contoh aneh merupakan indikasi stres padanya.

Sayangnya, kita tidak boleh mengabaikan bahwa programmer juga menggunakan kesadaran AI dalam serangan digital yang lebih maju dan lebih sulit dideteksi.

Teknologi terkini mengancam keamanan siber asosiasi. Tentu saja, bahkan dengan kemajuan baru dalam sistem pertahanan, kecakapan keamanan pada akhirnya akan gagal. Menggabungkan kualitas AI dengan kemampuan pakar keamanan dari pemeriksaan kelemahan hingga perlindungan ternyata sangat ampuh. Asosiasi mendapatkan sedikit pengetahuan, dengan demikian, mendapatkan waktu respons yang lebih singkat. AI untuk Keamanan Siber adalah gelombang baru dalam Keamanan yang ditunjukkan pada Gambar 3.2

Gambar 3.2 Ai Untuk Keamanan Siber.

Solusi Analisis Keamanan Siber AI Saat Ini untuk Perusahaan:

- Analisis Perspektif: Penentuan aktivitas yang diperlukan untuk investigasi atau reaksi.
- Analisis Diagnostik: Evaluasi pemeriksaan pendorong utama dan cara umum melakukan sesuatu dari episode dan serangan.
- Analisis Prediktif: Penentuan klien dan sumber daya dengan risiko lebih tinggi di kemudian hari dan kemungkinan bahaya yang akan datang [4].
- Analisis Detektif: Pengenalan bahaya yang tersembunyi, tidak jelas, bahaya yang terlewati, malware yang berkembang, dan perkembangan horizontal.
- Analisis Deskriptif: Untuk memperoleh status terkini dan pelaksanaan pengukuran dan pola.
- Pendekatan Manajemen Risiko yang Didukung AI untuk Keamanan Siber:
  - ✓ Pengumpulan Data yang Tepat.
  - ✓ Aplikasi Pembelajaran Representasi.
  - ✓ Kustomisasi Pembelajaran Mesin.
  - ✓ Analisis Ancaman Siber.
  - ✓ Model Masalah Keamanan

3.3 PENGGUNAAN KECERDASAN BUATAN DALAM KEAMANAN SIBER

- Dalam pengembangan, kerangka kerja keamanan digital berbasis AI dapat memberikan norma keamanan yang berhasil dan membantu mengembangkan teknik antisipasi dan pemulihan yang lebih baik. Di sisi lain, pemanfaatan AI untuk keamanan digital membantu menciptakan struktur verifikasi global yang dinamis, waktu nyata, yang mengubah akses area atau sistem ke manfaat.
- Faktanya, lebih dari 90 persen pakar keamanan siber AS dan Jepang mengantisipasi bahwa penyerang akan menggunakan AI untuk organisasi tempat mereka bekerja, sebagaimana ditunjukkan oleh pemeriksaan oleh Webroot.
- Latihan AI perusahaan memiliki berbagai macam kekurangan yang diantisipasi, termasuk penurunan nilai atau kontrol yang berbahaya terhadap data penyiapan, eksekusi, dan desain permainan bagian. Asosiasi keamanan siber Darktrace menyatakan bahwa kemajuan AI yang dimodifikasinya telah mendeteksi 63.500 ancaman gelap pada lebih dari 5.000 sistem, termasuk penyalahgunaan zero-day, risiko internal, dan serangan tersembunyi yang tidak terlihat [5].
- Dari keamanan sistem dan aplikasi web hingga perlindungan risiko dan akses terpadu dan aman, hal-hal keamanan canggih Fortune digunakan oleh sebagian besar asosiasi Fortune 500.
- Untuk asosiasi dan organisasi yang membutuhkan solusi keamanan digital, Spark Cognition menyediakan fitur AI yang mengidentifikasi dan melindungi dari malware, ransomware, Trojan horse, dan ancaman lainnya.
- Tampilan keamanan terkoordinasi Protector memberikan informasi tentang ancaman cloud bisnis atau lokasi hibrida.
- Intinya, instrumen tersebut dapat dengan cepat mengenali lalu lintas yang tidak biasa dalam sistem termasuk penambangan bitcoin, eksekusi file yang tidak dapat diakses, dan bahkan login dengan kekerasan – untuk memastikan keamanan seluruh asosiasi.
- Saat ini, AI merupakan inovasi terbaru untuk pemanfaatan AI secara praktis dalam keamanan siber.
- Produk keamanan siber mengumpulkan informasi dalam jumlah besar – penyelidik keamanan siber benar-benar menyerap informasi tersebut.
- AI menawarkan potensi yang sangat besar untuk membantu mengalahkan tantangan dan membantu asosiasi meningkatkan pola pikir keamanan siber mereka melalui investigasi kode yang cerdas dan pemeriksaan pengaturan serta pemeriksaan pergerakan.
- Meskipun AI digunakan di banyak bidang, keamanan siber adalah salah satu bidang yang telah mendapat perhatian khusus mengingat tingkat bahaya yang ditimbulkan dan jumlah serangan.
- Banyak administrator keamanan mengatakan bahwa mereka saat ini "sangat membutuhkan" inovasi AI untuk melindungi sistem dan informasi sensitif mereka.

- Namun, sistem keamanan digital berbasis udara dapat mengidentifikasi contoh perilaku berbahaya dalam lalu lintas jaringan dan data serta situs yang dikenal dengan sistem tersebut.
- Karena sistem keamanan berbasis kesadaran AI tidak bergantung pada tanda, sistem tersebut dapat mengidentifikasi serangan yang tidak praktis.
- Tahap federal umumnya terintegrasi ke dalam sistem perbankan atau bisnis dan dapat memperingatkan penipuan manusia dan analisis risiko jika benar-benar dianggap berisiko tinggi (berdasarkan faktor yang telah ditentukan sebelumnya), sehingga mempercepat prosedur identifikasi penipuan dan mengurangi hasil konstruktif yang salah.
- Organisasi tersebut menegaskan bahwa fondasinya dapat mendukung keamanan dan latihan operasional organisasi melalui pengakuan model pembelajaran terprogram dalam informasi pengaturan kronik. Barrier Storm mengatakan bahwa pengaturan SaaS mereka dapat memberikan fakultas keamanan TI di mengelola akun dengan akses ke informasi terkait acara di satu tempat melalui dasbor tunggal.
- DefenceStorm mengatakan telah mengoordinasikan jawaban investigasi SaaS-nya untuk memperbarui informasi dan kerangka kerja pemeriksaan Live Oak Bank terkini dalam beberapa bulan.
- Bank juga harus tahu bahwa upaya semacam itu oleh AI pada dasarnya direncanakan untuk mengumpulkan dan memilah informasi, sehingga menjamin bahwa informasi yang diidentifikasi dengan keamanan, misalnya, alamat IP, informasi firewall, dan kerangka kerja pencegahan gangguan, dikumpulkan dalam organisasi yang sebanding.
- Penelitian Kecerdasan Buatan Emery memungkinkan organisasi dan supervisor untuk bertahan dan menciptakan masalah AI yang bermasalah melalui Penelitian AI yang mendalam, panduan, dan potongan pengetahuan.
- Keamanan siber mengacu pada inovasi dan praktik yang dirancang untuk melindungi sistem dan data dari kerusakan atau akses yang tidak sah.
- Keamanan digital sangat penting mengingat badan pembuat undang-undang, organisasi, dan pasukan militer mengumpulkan metodologi dan menyimpan banyak informasi di komputer.
- Penyerang siber menginvestasikan uang dalam robotisasi untuk menangkal serangan, sementara banyak asosiasi masih menyelidiki upaya manual untuk menggabungkan hasil keamanan internal dan menempatkannya dalam pengaturan dengan data tentang bahaya eksternal.
- Sebagian besar pengaturan keamanan siber menggunakan pendekatan berbasis standar atau berbasis tanda tangan yang memerlukan lebih banyak mediasi manusia dan informasi kelembagaan.
- Pengetahuan buatan dapat meningkatkan efisiensi manusia sehingga dapat meningkatkan waktu yang dihabiskan untuk keamanan siber.

- Sejauh ini, pemerintah telah banyak menggabungkan kerangka kerja keamanan sibernya, yang telah menyebabkan cara yang berbeda untuk menangani kerangka kerja keamanan.
- Pemanfaatan AI dan sistem saraf AI telah memungkinkan para insinyur untuk menyesuaikan diri dengan vektor serangan baru dan meramalkan tahap-tahap selanjutnya dari para pelaku kejahatan dunia maya dengan lebih baik.
- Pemanfaatan AI juga dapat mendorong serangan-serangan komparatif dan mengarah pada periode baru serangan-serangan yang didukung negara dan pengintaian digital.
- Karena semakin banyak organisasi yang merangkul produk-produk berbasis AI dan AI sebagai komponen dari prosedur pertahanan mereka, para spesialis khawatir hal ini dapat menyebabkan keyakinan yang salah bahwa dunia ini baik-baik saja bagi para pekerja dan pakar TI.

3.4 PERAN AI DALAM KEAMANAN DUNIA MAYA

Penalaran AI (Kecerdasan buatan) dapat dicirikan sebagai dinamika buatan yang sebanding atau setara dengan dinamika manusia, berdasarkan perhitungan-perhitungan luar biasa tertentu dan perhitungan-perhitungan ilmiah terkait. Di sisi lain, Keamanan Dunia Maya mengacu pada upaya-upaya keamanan yang harus diambil untuk menangani serangan-serangan digital di dunia maya yang ditunjukkan pada Gambar 3.3.

Gambar 3.3 Peran Kecerdasan Buatan Dalam Keamanan Siber.

Untuk keamanan siber, kecerdasan buatan dapat memecah sejumlah besar informasi, membantu kerangka kerja dan pemrograman yang tepat untuk menentukan pilihan dan memperoleh penurunan yang luar biasa dalam serangan dan inkonsistensi dengan cara yang jauh lebih cepat. Karena dapat bekerja 24/7 tanpa istirahat, ia lebih unggul daripada pekerja manusia. Kecerdasan berbasis komputer akan memungkinkan pengujian pemrograman

terkomputerisasi untuk menemukan dan menghancurkan bug sebelum mereka muncul untuk menjaga jarak strategis dari setiap peluang finansial pada klausul pelarian.

Kecerdasan Simulasi Dapat Membedakan Serangan Digital

Kecerdasan simulasi dapat digunakan secara efektif untuk mengenali serangan tak terduga di internet, termasuk berbagai tahap web dan situs otoritas berbasis keamanan tinggi. Programmer menggunakan berbagai pendekatan untuk memulai serangan digital dan minat untuk pulih. Dalam situasi ini, situs yang membutuhkan keamanan tinggi bergantung pada kecerdasan simulasi sebagai strategi penting untuk mengenali serangan digital. Selain itu, sulit bagi programmer untuk mendapatkan akses ke situs dengan keamanan tinggi karena situs dengan keamanan tinggi bergantung pada kecerdasan buatan untuk mengidentifikasi bagian yang tidak disetujui tersebut. Laju pencapaian yang tinggi dari situs berbasis kecerdasan buatan dalam mengidentifikasi serangan digital menyebabkan situs sejenis memilih kecerdasan berbasis komputer sebagai upaya keamanan yang penting.

Kecerdasan Berbasis Komputer Dapat Mencegah Serangan Digital

Kita dapat melihat bahwa bukti yang dapat dikenali yang tidak penting dari bahaya keamanan tidak dapat membantu situs panggung virtual untuk menghindari penyerang digital termasuk programmer. Dalam situasi ini, kecerdasan buatan dapat digunakan untuk mencegah serangan digital dengan berbagai cara. Untuk mencegah serangan digital, individu yang bertanggung jawab atas situs harus berpikir seperti programmer. Di sini, kecerdasan berbasis komputer dapat memanfaatkan cara programmer berpikir dan bertindak untuk memecahkan kode keamanan.

Kecerdasan Buatan dan Keamanan Siber Cakupan Luas

Bayangkan situs dengan lalu lintas yang lebih sedikit, beberapa kerangka kerja yang saling terhubung, dan kurang menonjol. Tidak ada gunanya mengandalkan kecerdasan buatan yang kompleks untuk melindungi situs dari serangan digital. Selain itu, programmer mungkin tidak menargetkan situs yang kurang dikenal dan panggung terbuka karena mereka tidak dapat meningkatkan banyak dari serangan yang diusulkan. Mengembangkan inovasi menempatkan keamanan siber pada risiko yang serius. Faktanya, bahkan tingkat kemajuan baru dalam kerangka kerja pertahanan spesialis keamanan pun akhirnya meleset.

Demikian pula, karena sistem pertahanan dan kemajuan yang mengancam berjalan dalam siklus tanpa akhir, sifat dan volume serangan siber yang beraneka ragam telah meluas. Dengan memperkuat sifat kesadaran Arteri dengan keamanan siber, spesialis keamanan memiliki sumber daya tambahan untuk mempertahankan kerangka kerja dan data yang rapuh dari penyerang terkomputerisasi. Setelah menerapkan inovasi ini, inovasi ini membawa sedikit pengetahuan, yang mengakibatkan berkurangnya waktu reaksi. Cap Gemini baru-baru ini merilis laporan yang bergantung pada kecerdasan simulasi dalam keamanan siber, yang menyatakan bahwa 42% organisasi yang dipertimbangkan telah melihat peningkatan dalam episode keamanan melalui aplikasi yang peka waktu.

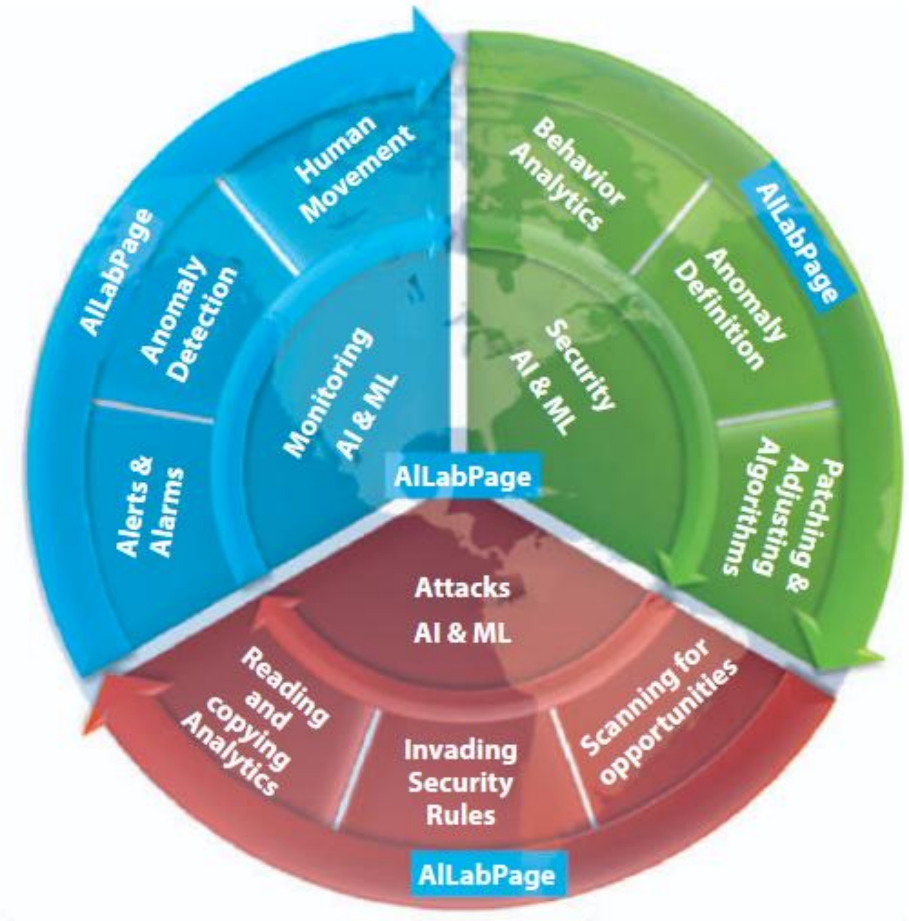
Laporan itu juga menemukan bahwa dua dari tiga asosiasi ingin merangkul pengaturan kecerdasan buatan pada tahun 2020. Keamanan data saat ini menjadi masalah yang lebih besar dari sebelumnya. Menghidupkan kembali rencana permainan keamanan siber yang ada

dan mempertahankan setiap lapisan keamanan yang mungkin berlaku tidak menjamin bahwa data Anda aman. Bagaimanapun, memiliki bantuan yang kuat dari kemajuan mutakhir akan mendorong tugas spesialis keamanan.

Tantangan dan Janji Kecerdasan Buatan dalam Keamanan Siber

Keamanan siber bukan hanya divisi inovasi data atau masalah yang melibatkan individu di kantor yang sama. Ini adalah aktivitas setiap pekerja dan bahkan klien di kota. Sementara spesialis keamanan siber telah mengakui kecerdasan buatan sebagai nasib akhir bisnis, menemukan jawaban untuk masalahnya belum cukup ditangani. Selain sebagai jawaban, ini adalah bahaya yang signifikan bagi organisasi. Penalaran AI dapat secara efektif mengeksplorasi praktik pelanggan, menyelesaikan model, dan memahami berbagai variasi dari standar atau variasi dari norma dalam kerangka kerja.

Dengan data tersebut, jauh lebih mudah untuk memahami kekurangan canggih dengan cepat. Di sisi lain, kewajiban yang sekarang bergantung pada data manusia pada saat itu akan menjadi lemah terhadap usaha komputerisasi yang mengancam yang meniru komputasi berbasis wawasan yang ditiru secara nyata. Beberapa organisasi sedang terburu-buru untuk mengeluarkan barang-barang berbasis AI mereka di pasar.



Gambar 3.4 Tantangan Dalam Keamanan Siber.

Dengan petunjuk ini, mereka mungkin mengabaikan keseriusan situasi, menarik kesimpulan yang salah bahwa semuanya baik-baik saja di dunia. Bergantung pada "pembelajaran terarah" adalah ancaman lain. Di bawah ini, estimasi menamai bermacam-macam pendidikan sebagaimana ditunjukkan oleh perilakunya. Itu bisa berupa malware, data bersih, atau tag lainnya. Penjahat dunia maya, jika mereka memperoleh akses ke perusahaan keamanan, dapat mengubah nama untuk keuntungan mereka. Selain itu, upaya rutin yang mengandalkan pengetahuan berbasis PC dapat dibatasi oleh upaya peretasan mutakhir yang memanfaatkan kecerdasan buatan.

Terlepas dari menjadi bahaya keamanan bagi asosiasi, penalaran terkomputerisasi akan terus membatasi komitmen keamanan yang khas dengan hasil yang terbaik. Otomatisasi penalaran terkomputerisasi akan memiliki pilihan untuk mengenali kejadian yang berulang dan bahkan memperbaikinya. Itu juga akan memiliki alternatif untuk mengarahkan ancaman orang dalam dan perangkat yang dijelaskan oleh para eksekutif pada Gambar 3.4 di atas.

Keamanan Siber Masa Kini dan Masa Depannya dengan Kecerdasan Simulasi

Saat ini, bisnis dan organisasi lain mempertimbangkan keamanan kerangka kerja mereka dengan saksama. Mereka memikirkan dampak kolosal dari setiap serangan komputerisasi yang kecil hingga besar. Untuk memastikan terhadap serangan tersebut, mereka menggunakan berbagai lini penjaga gerbang. Struktur keamanan berlapis ini umumnya dimulai dengan firewall yang paling sesuai untuk mengendalikan dan menyaring lalu lintas kerangka kerja. Setelah lapisan ini, lini pertahanan kedua mencakup program antivirus (Pemrograman AV). Instrumen AV ini memeriksa sistem untuk menemukan dan menghapus kode dan catatan yang berbahaya. Dengan dua lini pertahanan ini, organisasi secara andal menjalankan benteng sebagai bagian dari rencana pemulihan bencana.

Saat ini, menyiapkan strategi firewall, mengawasi bala bantuan, dan berbagai tugas semacam itu memerlukan seorang ahli, tetapi kecerdasan buatan akan mengubah metodologi konvensional.

- Organisasi akan memiliki opsi untuk menyaring dan menanggapi kejadian keamanan dengan memanfaatkan peralatan canggih.
- Firewall mutakhir akan memiliki inovasi AI bawaan yang dapat menemukan contoh dalam kumpulan jaringan dan mengurutkannya secara otomatis setiap kali dianggap sebagai bahaya.
- Dapat diprediksi, kemampuan bahasa karakteristik kecerdasan buatan akan digunakan untuk memahami awal serangan digital. Hipotesis ini dapat digabungkan dengan memeriksa informasi melalui web.

Keamanan Siber yang Lebih Baik dengan Kecerdasan Berbasis Komputer dan AI (ML)

Metode peretasan yang rumit, misalnya, membingungkan, polimorfisme, dan lainnya, menjadikannya pengujian yang dapat diandalkan untuk memahami aktivitas berbahaya. Selain itu, insinyur keamanan dengan kekurangan tenaga kerja ekspres wilayah merupakan masalah lain. Dengan penalaran buatan manusia yang merambah ke keamanan siber, otoritas dan pemeriksa berusaha menggunakan kapasitasnya untuk mengenali dan memeriksa penyergapan canggih yang canggih dengan intervensi manusia yang tidak relevan. Kerangka

kerja penalaran kecerdasan buatan dan AI, bagian dari wawasan berbasis komputer, telah melibatkan spesialis keamanan untuk mendapatkan beberapa jawaban mengenai vektor serangan baru.

AI dalam keamanan siber jauh lebih dari sekadar pemanfaatan kalkulasi yang dapat diabaikan. AI cenderung digunakan untuk memecah bahaya digital dengan lebih baik dan bereaksi terhadap episode keamanan. Ada beberapa keuntungan besar AI lainnya, sebagai berikut:

- Mendeteksi aktivitas ganas dan menghentikan serangan digital
- Menganalisis titik akhir seluler untuk bahaya digital; Google saat ini menggunakan AI untuk hal yang sama
- Meningkatkan pemeriksaan manusia; dari lokasi serangan berbahaya hingga asuransi titik akhir
- Penggunaan dalam mengotomatiskan tugas keamanan sehari-hari
- Tidak ada kelemahan zero-day

Pengadopsi AI Bergerak untuk Melakukan Langkah

Pengetahuan berbasis komputer baru-baru ini dipahami untuk memperkuat sistem keamanan organisasi. Ada berbagai model nyata di mana tindakan yang dikendalikan oleh penalaran AI secara umum meningkatkan keamanan siber.

- ➲ Gmail menggunakan AI untuk memblokir seratus juta spam dalam sehari. Gmail telah mengembangkan sistem untuk menyaring pesan dan menawarkan kondisi bebas spam secara menguntungkan.
- ➲ Persiapan ilmiah Watson IBM menggunakan AI untuk mengenali bahaya tingkat lanjut dan rencana keamanan siber lainnya.
- ➲ Google menggunakan wawasan buatan Pembelajaran Mendalam pada tahap Pengetahuan Video Cloud-nya. Pada tahap ini, kronik yang disisihkan pada pekerja dieksplorasi bergantung pada substansi dan pengaturannya. Perhitungan pengetahuan yang direkayasa mengirimkan peringatan keamanan setiap kali sesuatu yang mencurigakan ditemukan.
- ➲ Tahap Belbin menggunakan keinginan risiko yang dikendalikan kesadaran AI untuk menjamin pendirian TI terhadap data dan entri keamanan.

3.5 DAMPAK AI PADA KEAMANAN SIBER

Saat ini, ada diskusi besar yang sedang berlangsung tentang apakah penalaran terkomputerisasi (Kecerdasan buatan) merupakan hal yang menguntungkan atau tidak menguntungkan sejauh menyangkut dampaknya terhadap kehidupan manusia. Dengan semakin banyaknya upaya yang memanfaatkan kecerdasan buatan untuk kebutuhan mereka, inilah saat yang tepat untuk menguraikan dampak potensial dari penerapan kecerdasan buatan dalam bidang keamanan digital yang ditunjukkan pada Gambar 3.5.

Gambar 3.5 Dampak Ai Dalam Keamanan Siber

1. **Waktu Penemuan dan Reaksi yang Lebih Cepat:**
   Penalaran terkomputerisasi dapat mempercepat pengakuan masalah yang sebenarnya, dengan cepat merujuk silang berbagai peringatan dan sumber data keamanan. Profesional keamanan terkomputerisasi manusia bahkan akan menjadikan metodologi sebagai kebutuhan situasi yang harus ditangani. Namun, hal ini cenderung lebih terbantu oleh struktur kecerdasan simulasi yang menyarankan rencana untuk meningkatkan respons.
2. **Keamanan Sistem:**
   Dua bagian penting dari keamanan sistem adalah pembuatan prosedur keamanan dan pemahaman geologi sistem organisasi. Biasanya, kedua aktivitas ini sangat berulang. Kita dapat menggunakan kecerdasan buatan untuk mempercepat teknik ini, yang dilakukannya dengan mengamati dan mempelajari struktur lalu lintas sistem serta menyarankan tindakan keamanan. Itu tidak hanya menghemat waktu, tetapi juga menghemat banyak tenaga dan sumber daya yang dapat kita terapkan pada area pengembangan dan kemajuan baru yang mekanis.
3. **Penemuan dan Pengendalian Penanggulangan Phishing:**
   Salah satu strategi serangan komputer yang paling umum digunakan, di mana para insinyur perangkat lunak berusaha untuk meneruskan muatan mereka menggunakan serangan phishing, adalah phishing. Pesan phishing adalah hal yang umum; satu dari setiap 99 pesan adalah serangan phishing. Untungnya, kecerdasan berbasis komputer ML dapat menerima aktivitas penting dalam mencegah dan mengalihkan serangan phishing.
   Kecerdasan berbasis komputer ML dapat mengenali dan melacak lebih dari sepuluh ribu otoritas phishing dinamis dan merespons serta memulihkan secara signifikan lebih cepat daripada yang dapat dilakukan oleh individu. Selain itu, simulasi kecerdasan buatan (ML) berfungsi untuk memisahkan risiko phishing dari seluruh dunia. Tidak ada batasan dalam pengenalan upaya phishing terhadap wilayah atau area tertentu. Pengetahuan berbasis komputer memungkinkan untuk membedakan antara situs palsu dan situs asli dengan cepat.

4. **Validasi Aman:**

   Kata sandi secara andal sangat rapuh dalam hal keamanan. Selain itu, kata sandi sering kali menjadi penghalang utama antara penjahat digital dan catatan kita. Cara utama konfirmasi aman dapat dikembangkan adalah dengan bukti fisik yang jelas, di mana kecerdasan buatan menggunakan berbagai komponen untuk mengenali seseorang. Misalnya, nirkabel dapat menggunakan pemindai sidik jari dan afirmasi wajah yang luar biasa untuk memungkinkan Anda masuk. Strategi di balik ini mencakup program yang melihat data penting yang berpusat pada wajah dan jari Anda untuk melihat apakah login tersebut valid. Selain itu, simulasi kecerdasan buatan dapat memeriksa berbagai segmen untuk memilih apakah pelanggan tertentu diizinkan untuk masuk ke gadget tertentu. Teknisi menganalisis faktor-faktor seperti cara Anda memasukkan tombol, kecepatan menulis, dan tingkat kesalahan saat menulis.

Penggunaan Positif Berbasis AI untuk Keamanan Siber

- ☑ Login biometrik secara logis digunakan untuk membuat login yang aman dengan melihat sidik jari, data, atau telapak tangan. Ini dapat digunakan sendiri atau diidentifikasi dengan ekspresi misteri dan mulai digunakan di sebagian besar PDA baru. Asosiasi besar telah menjadi penanggulangan entri keamanan yang membahayakan alamat email, informasi tunggal, dan kata sandi. Otoritas keamanan terkomputerisasi telah menekankan berbagai kesempatan bahwa kata sandi lemah terhadap serangan kubus, pertukaran informasi tunggal, informasi kartu kredit, dan nomor dana investasi yang diawasi pemerintah. Ini sebagian besar merupakan alasan mengapa login biometrik merupakan komitmen kecerdasan buatan yang positif terhadap keamanan digital.
- ☑ Informasi buatan juga dapat digunakan untuk mengisolasi risiko dan aktivitas berbahaya lainnya. Struktur standar tidak dapat terus memantau jumlah malware yang dibuat setiap bulan, jadi ini adalah lokasi yang memungkinkan bagi pemahaman berbasis PC untuk masuk dan menemukan masalah ini. Organisasi keamanan siber melatih kerangka kerja intelijen berbasis komputer untuk membedakan infeksi dan malware dengan memanfaatkan kalkulasi yang rumit sehingga intelijen simulasi kemudian dapat menjalankan pengenalan desain dalam pemrograman. Kerangka kerja intelijen berbasis komputer dapat disiapkan untuk mengenali bahkan tindakan serangan ransomware dan malware yang paling kecil sebelum mereka memasuki struktur dan beberapa saat kemudian mengisolasinya dari sistem itu. Mereka juga dapat menggunakan batasan bijaksana yang mengalahkan kecepatan strategi konvensional.
- ☑ Struktur yang tiba-tiba meningkatkan permintaan akan intelijen berbasis komputer membuka kemampuan untuk pelatihan bahasa umum yang mengumpulkan data secara alami dengan membaca artikel, berita, dan studi tentang ancaman digital. Potensi untuk pelatihan bahasa umum yang mengumpulkan informasi secara alami dengan membaca artikel, berita, dan studi tentang ancaman tingkat lanjut. Informasi

ini dapat memberikan pemahaman tentang penyimpangan, penyergapan terkomputerisasi, dan teknik ekspektasi. Hal ini memungkinkan perusahaan keamanan terkomputerisasi untuk tetap waspada terhadap bahaya dan kerangka waktu terbaru serta membuat strategi responsif untuk menjaga keamanan organisasi.

- ☑ Kerangka kerja intelijen berbasis komputer juga dapat digunakan dalam situasi verifikasi multifaset untuk memberikan akses kepada klien mereka. Klien yang berbeda dari suatu organisasi memiliki berbagai tingkat manfaat verifikasi yang juga bergantung pada area tempat mereka mendapatkan informasi. Pada saat intelijen berbasis komputer digunakan, sistem validasi dapat menjadi jauh lebih kuat dan berkelanjutan dan dapat mengubah manfaat akses tergantung pada sistem dan area klien. Verifikasi multifaset mengumpulkan data klien untuk memahami perilaku individu ini dan membuat jaminan tentang manfaat akses klien.
- ☑ Untuk menggunakan kecerdasan buatan secara maksimal, sangat penting bahwa hal itu diselesaikan oleh perusahaan keamanan canggih yang peduli dengan cara kerjanya. Meskipun sebelumnya, penyergapan malware dapat terjadi tanpa meninggalkan tanda-tanda kelemahan yang disalahgunakan, pengetahuan yang direproduksi dapat masuk untuk menjamin perusahaan keamanan tingkat lanjut dan klien mereka dari serangan apa pun, ketika ada berbagai penyergapan berbakat yang terjadi [16].

Kelemahan dan Pembatasan Penggunaan Penalaran Terkomputerisasi untuk Keamanan Digital

- ❖ Keuntungan yang diuraikan di atas hanyalah sebagian kecil dari kapasitas kesadaran AI dalam mendukung keamanan terkomputerisasi; ada juga keharusan yang membuat wawasan berbasis komputer tidak berubah menjadi instrumen standar yang digunakan di lapangan. Untuk membuat dan memelihara sistem penalaran terkomputerisasi, asosiasi akan membutuhkan sebagian besar keuntungan termasuk memori, data, dan daya penanganan. Selain itu, mengingat fakta bahwa sistem pengetahuan yang diciptakan kembali disiapkan melalui pembelajaran file pendidikan, perusahaan keamanan terkomputerisasi perlu mendapatkan berbagai macam file informatif kode malware, kode tidak beracun, dan keanehan.
- ❖ Pengadaan koleksi pendidikan yang tepat ini dapat memakan waktu dan sumber daya yang sangat lama yang tidak dapat dikelola oleh beberapa asosiasi.
- ❖ Kerugian: Kerugian lainnya adalah bahwa para insinyur perangkat lunak dapat dengan cara yang sama menggunakan pengetahuan buatan sendiri untuk menguji malware mereka dan mengembangkan serta merombaknya hingga mungkin menjadi anti-kecerdasan buatan. Sejujurnya, malware penegasan pengetahuan yang diciptakan kembali bisa sangat berbahaya karena mereka dapat mengambil alih dari gadget wawasan buatan yang ada dan mengembangkan serangan yang diciptakan lebih lanjut untuk memiliki alternatif untuk memasuki program keamanan terkomputerisasi biasa atau bahkan sistem yang dibantu wawasan berbasis PC.

Solusi untuk Pembatasan Kecerdasan Buatan

Mengetahui pembatasan dan kerugian ini, jelas wawasan berbasis PC masih jauh dari berubah menjadi tindakan keamanan digital yang mendasar. Pendekatan terbaik untuk sementara waktu adalah mendapatkan prosedur biasa bersama dengan instrumen wawasan yang direproduksi, sehingga organisasi harus mengingat tindakan ini saat menyusun strategi keamanan terkomputerisasi mereka:

- Mempekerjakan perusahaan keamanan siber dengan spesialis yang memiliki pemahaman dan kapasitas dalam berbagai fitur keamanan digital.
- Mintalah tim keamanan komputer Anda menguji struktur dan kerangka kerja Anda untuk setiap celah potensial dan segera memperbaikinya.
- Gunakan saluran untuk URL guna memblokir asosiasi berbahaya yang mungkin memiliki penyakit atau malware.
- Pasang firewall dan pemindai malware lainnya untuk memastikan sistem Anda aman dan selalu perbarui untuk mengatur malware terbaru.
- Pantau lalu lintas dinamis Anda dan terapkan saluran keluar untuk membatasi lalu lintas tersebut.
- Tinjau risiko komputer dan tampilan keamanan terbaru secara terus-menerus untuk mendapatkan informasi tentang bahaya mana yang harus Anda tangani terlebih dahulu dan kembangkan tampilan keamanan Anda dengan cara yang sama.
- Lakukan audit rutin terhadap peralatan dan perangkat lunak untuk memastikan struktur Anda kuat dan berfungsi.

Mengikuti metode ini dapat membantu mengurangi sejumlah besar ancaman yang terkait dengan serangan canggih, tetapi penting untuk dipahami bahwa organisasi Anda masih berisiko mengalami serangan. Oleh karena itu, keseimbangan saja tidak cukup dan Anda juga harus bekerja sama dengan istilah keamanan canggih Anda untuk mengembangkan teknik pemulihan. Karena kapasitas kecerdasan buatan diteliti untuk membantu profil keamanan komputer suatu organisasi, kecerdasan buatan juga dibuat oleh para insinyur perangkat lunak. Bagaimanapun, kecerdasan buatan masih dalam tahap pengembangan dan batas inertnya masih jauh dari jangkauan, kita belum dapat mengetahui apakah kecerdasan buatan suatu hari nanti akan berguna atau tidak bagi keamanan komputer. Sementara itu, penting bagi organisasi untuk melakukan sebanyak mungkin dengan memadukan strategi tradisional dan pengetahuan berbasis komputer agar tetap konsisten dengan strategi keamanan canggih mereka.

3.6 ANCAMAN KEAMANAN KECERDASAN BUATAN

Semua serangan siber dapat diklasifikasikan dalam istilah aksesibilitas dan kepercayaan yang paling dikenal luas, yang saling terkait untuk membentuk tiga arah utama yang ditunjukkan pada Gambar 3.6. Aktivitas Spionase atau Rahasia, yang berkaitan dengan keamanan siber menyiratkan pengumpulan pengalaman tentang kerangka kerja dan menggunakan data yang diperoleh untuk keuntungan mereka sendiri atau merencanakan

serangan yang lebih maju. Pada akhirnya, seorang programmer dapat menggunakan motor berbasis ML untuk menembus dan menjadi terbiasa dengan konten seperti kumpulan data.

Gambar 3.6 Ancaman Keamanan.

Sabotase (Kerusakan) dengan tekad untuk melumpuhkan kegunaan kerangka kerja intelijen berbasis komputer dengan membanjiri kelincahan simulasi dengan permintaan, atau perubahan model. Penipuan (Misrepresentasi) yang dalam kecerdasan buatan menandakan kesalahan klasifikasi tugas, misalnya, menyajikan informasi yang tidak tepat dalam kumpulan data persiapan (kerusakan data) atau menghubungkan dengan kerangka kerja pada tahap pengetahuan atau konsepsi.

Memperluas Ancaman Keamanan Siber dengan Kesadaran Buatan

Setiap hari, orang-orang berupaya mengubah cara masyarakat menghadapi kemajuan, dan salah satu perkembangan terbaru di bidang rekayasa perangkat lunak adalah AI. Berbagai organisasi meneliti penggunaan kecerdasan berbasis komputer dan AI untuk melihat cara mengamankan kerangka kerja mereka terhadap serangan digital dan malware. Namun, mengingat kecenderungannya untuk belajar sendiri, sistem kecerdasan buatan ini kini juga telah mencapai tingkat di mana mereka dapat dipersiapkan untuk menjadi bahaya bagi struktur, yaitu, masuk ke mode serangan keras dan cepat.

Tidak dapat dihindari bahwa kita akan melihat peningkatan penggunaan kecerdasan buatan dalam kehidupan sehari-hari. Namun, seperti berbagai perkembangan dalam pasar teknologi, hal ini juga membuka jalan yang sama sekali berbeda untuk penyalahgunaan yang dikembangkan oleh para penyerang digital. Ada kemungkinan yang jelas untuk tingkat penyalahgunaan yang melampaui apa yang telah kita lihat sejauh ini, terutama karena teknologi terus maju. Mari kita bahas bagaimana AI memengaruhi keamanan siber secara negatif.

1. Peretas memperoleh keunggulan dengan kecerdasan buatan:

Ahli keamanan siber mengakui bahwa pengenalan kekuatan otak buatan manusia juga penting bagi penipu dan programmer. Penjahat siber dapat menggunakan komputerisasi untuk menandai arah penemuan kelemahan baru yang dapat mereka manfaatkan dengan cepat dan

menciptakan beberapa masalah. Ilmuwan dan pakar khawatir tentang ancaman yang ditimbulkan oleh model inovasi kecerdasan buatan terhadap keamanan siber, yang sebagian besar menjaga PC dan data kita dan PC serta data organisasi dan pemerintah — aman dari penjahat siber. Ketakutan sebagian orang adalah bahwa kecerdasan berbasis komputer akan membawa serta dimulainya berbagai jenis bahaya canggih baru yang menghindari teknik normal untuk melawan serangan. Misalnya, AI dapat melakukan serangan terhadap konten dengan kecepatan dan tingkat multifaset yang umumnya tidak dapat ditandingi oleh individu. Berikut adalah bahaya lebih lanjut terhadap keamanan digital melalui kecerdasan berbasis komputer.

2. Menjadikan robot dan kendaraan sebagai senjata:

Para ahli dari Cambridge dan Oxford mengantisipasi bahwa kecerdasan buatan dapat dilatih untuk meretas kendaraan dan robot yang dapat mengemudi sendiri, sehingga menciptakan kemungkinan terjadinya kecelakaan kendaraan yang disengaja dan pengeboman yang tidak terkendali. Misalnya, kendaraan independen Waylon milik Google menerapkan pembelajaran mendalam, dan kerangka kerja tersebut dapat diretas untuk menganggap tanda berhenti sebagai lampu hijau, yang dapat menyebabkan kecelakaan fatal. Lebih jauh, seiring meningkatnya kerumitan kecerdasan buatan dan penjahat dunia maya, demikian pula kemungkinan terjadinya serangan terus-menerus. Misalnya, beberapa programmer dapat menggabungkan robot ke dalam sekumpulan robot yang dapat diisi dengan bahan peledak untuk menyebabkan serangan mematikan. Inovasi penalaran terkomputerisasilah yang melibatkan penjahat dunia maya untuk memprogram serangan ini dengan lebih efektif dan mengaitkan robot dengan informasi penting.

3. Programmer Bot:

Kita menghargai berbicara dengan chatbot tanpa menyadari jumlah informasi yang kita berikan kepada mereka. Demikian pula, chatbot dapat dimodifikasi untuk menjaga percakapan dengan klien dengan cara yang akan membujuk mereka untuk mengungkapkan data moneter atau pribadi mereka, asosiasi, dan sebagainya. Pada tahun 2016, bot Facebook menyebut dirinya sebagai pendamping dan menipu 10.000 klien Facebook ke dalam instalasi malware. Setelah malware tersebut terancam, ia menyimpan catatan Facebook klien tersebut. Botnet yang mendukung kecerdasan buatan dapat melemahkan SDM melalui telepon dan dukungan online.

Sebagian besar dari kita menggunakan bot percakapan kecerdasan berbasis komputer, misalnya, Alexa Amazon atau Google Aide tetapi kita tidak memahami sejauh mana informasi yang mereka miliki tentang kita. Menjadi inovasi yang digerakkan oleh IoT, mereka sebagian besar dapat mendengar bahkan percakapan pribadi yang terjadi di sekitar mereka. Selain itu, beberapa chatbot tidak diatur secara efektif untuk transmisi data yang aman, misalnya, Verifikasi Tingkat Transportasi (TLV) atau konvensi HTTPS dapat digunakan secara memadai oleh programmer.

- Spear-phishing menjadi lebih mudah:
  Kesadaran buatan dalam serangan keamanan juga akan memudahkan penyerang digital tingkat rendah untuk mengendalikan gangguan kompleks hanya dengan memprosesnya dengan mudah. Pengembang secara rutin menang dengan meningkatkan skala tugas mereka. Semakin banyak orang yang mereka ikuti rencana phishing atau, semakin banyak kerangka kerja yang mereka selidiki, hampir dapat dipastikan mereka akan sampai ke tempat yang mereka inginkan. AI melengkapi mereka dengan cara untuk menangani skala ke tingkat yang jauh lebih tinggi, melalui mekanisasi sasaran dan menyampaikan serangan massal. Titik acuan utama di mana penjahat dunia maya menggunakan kecerdasan berbasis komputer untuk melancarkan serangan adalah melalui lance phishing. Sistem penalaran terkomputerisasi dengan bantuan model AI dapat menyelamatkan diri dari banyak masalah dengan meniru individu dengan membuat pesan palsu yang memengaruhi. Menerapkan metode ini, programmer dapat menggunakannya untuk melakukan lebih banyak serangan phishing. Programmer juga dapat menggunakan kecerdasan berbasis komputer untuk membuat malware untuk proyek yang menyesatkan atau kotak pasir yang mencoba menemukan kode pemberontak sebelum dikirim dalam pengaturan asosiasi.
  Selain itu, trik phishing berbasis kecerdasan buatan hanyalah asal-usulnya. Dengan menggunakan AI, penyerang digital dapat mencari kelemahan potensial dan mengotomatiskan cakupan korban potensial mereka. Inovasi yang sebanding dapat membantu mereka dalam membedah sistem perlindungan digital berbasis kecerdasan buatan secara memadai dan menghasilkan jenis malware baru yang dapat menyelinap melaluinya.
- Pencemaran berbahaya
  Latihan kecerdasan buatan asosiasi menghadirkan berbagai kelemahan yang diharapkan, menggabungkan kerusakan jahat atau informasi pemrosesan, pemanfaatan, dan pengaturan porsi. Tidak ada industri yang aman, dan ada berbagai pengelompokan di mana kecerdasan buatan dan AI mulai sekarang memiliki tugas dan dengan demikian menghadirkan bahaya yang berkepanjangan. Misalnya, trik Visa mungkin menjadi mudah. Selain itu, kerangka kerja keamanan, kondisi, dan kesejahteraan mungkin dipertaruhkan yang mengendalikan perangkat fisik digital yang mengawasi pengarahan kereta api, arus lalu lintas, atau bendungan.
- Perencanaan sistem sosial
  Bahaya lain yang berbasis kecerdasan buatan akan menggabungkan perencanaan komunikasi interpersonal jarak jauh tingkat tinggi. Misalnya, perangkat yang didukung kecerdasan berbasis komputer yang akan melihat lebih jauh ke dalam tahap komunikasi interpersonal jarak jauh akan memungkinkan militan psikologis untuk mengidentifikasi kota dan target manusia yang tepat dan bekerja lebih efektif. Dengan cara ini, pakar keamanan siber dan kantor perlawanan harus bekerja sama untuk memahami bahaya tersebut dan membuat pengaturan.

- Serangan rumah

  Berbagai bagian dari kehidupan kita sendiri sudah dirobotisasi dengan perangkat terkait IoT dan pembantu jarak jauh. Namun, ini membutuhkan banyak data pribadi yang ada di cloud. Pengaturan asosiasi yang rumit ini akan menciptakan elemen kelemahan baru, dengan bahaya digital yang menyerang lebih dekat ke rumah misalnya, penindas berbasis rasa takut, pemerintah yang tidak patuh, dan programmer dapat menargetkan perangkat klinis yang terhubung ke web. Beberapa waktu dari sekarang, kerangka kerja peringatan dan kunci mungkin tidak cukup untuk melindungi kita di rumah. Bot yang tidak patuh dapat memeriksa kerangka kerja, memburu kelemahan. Dengan demikian, sasaran yang menarik adalah sasaran yang rentan. Hal ini menyiratkan bahwa siapa pun dapat berubah menjadi sasaran potensial.

3.7 PEMANFAATAN AI DALAM KEAMANAN SIBER

Kecerdasan buatan telah membuat beberapa kemajuan dalam bidang keamanan siber dan beberapa pedagang kecerdasan buatan telah meyakinkan untuk mendorong hal-hal yang menggunakan wawasan buatan untuk membantu melindungi terhadap ancaman terkomputerisasi. Di Emerj, kami telah melihat berbagai pedagang keamanan siber yang menawarkan pengetahuan berbasis komputer dan hal-hal berbasis kecerdasan buatan untuk membantu memahami dan mengelola ancaman tingkat lanjut. Di Amerika Serikat, Pentagon menciptakan Pusat Kecerdasan Buatan Gabungan (JAIC) untuk membantu melindungi infrastruktur penting AS dari aktivitas siber yang berbahaya. Dalam artikel ini, kami membahas beberapa kasus penggunaan kecerdasan buatan yang lebih umum dalam keamanan siber, yang telah dibuktikan dengan beberapa bukti penggunaan bisnis yang sebenarnya. Secara khusus, kami membahas:

AI untuk Bukti Pembeda Bahaya Sistem

- ✓ Pemeriksaan Email AI
- ✓ Pemrograman Antivirus Berbasis AI
- ✓ Demonstrasi Perilaku Klien Berbasis AI
- ✓ AI untuk Memerangi Ancaman Berbasis AI

Kami memulai pemeriksaan kecerdasan berbasis komputer dalam ruang keamanan siber dengan klarifikasi mengapa kecerdasan simulasi sangat cocok untuk keamanan siber.

Kecocokan Umum untuk Kesadaran Buatan dalam Keamanan Siber

Bagi bisnis yang melindungi data mereka, mengatur keamanan sangatlah penting, dan bahkan bisnis kecil mungkin memiliki banyak aplikasi yang berjalan, yang semuanya memerlukan pendekatan keamanan yang berbeda yang disetujui. Profesional manusia mungkin memerlukan beberapa hari hingga minggu untuk benar-benar menghargai rencana permainan ini dan menjamin penerapan keamanan yang baik. Keamanan siber pada hakikatnya mencakup kebosanan dan kejenuhan. Hal ini karena membedakan bukti dan evaluasi ancaman siber memerlukan penelusuran melalui sejumlah besar informasi dan pencarian titik-titik informasi yang aneh.

Organisasi dapat memanfaatkan informasi yang dikumpulkan oleh program keamanan sistem berbasis standar mereka saat ini untuk menyiapkan kalkulasi intelijen berbasis komputer guna mengenali ancaman siber baru. Memahami hasil serangan dan reaksi yang diperlukan dari organisasi juga memerlukan pemeriksaan informasi lebih lanjut. Kalkulasi kecerdasan buatan dapat disiapkan untuk membuat langkah-langkah tertentu yang telah ditentukan sebelumnya jika terjadi serangan dan setelah beberapa waktu dapat menyadari apa reaksi terbaik yang seharusnya melalui kontribusi dari spesialis topik keamanan siber.

Spesialis keamanan manusia tidak dapat menandingi kecepatan dan skala di mana program intelijen berbasis komputer dapat mencapai tugas investigasi informasi ini. Lebih jauh, program investigasi informasi keamanan siber berbasis kecerdasan buatan dapat menyelesaikan tugas dengan presisi yang jauh lebih tinggi daripada pakar manusia. Pemeriksaan informasi skala besar dan pengenalan penyimpangan adalah sebagian dari area di mana kecerdasan buatan dapat menambah nilai saat ini dalam keamanan siber. Banyak gangguan keamanan siber biasanya bekerja selama pengaturan proyek mengamati informasi yang mengalir melalui sistem adalah salah satu pendekatan untuk mengidentifikasi bahaya keamanan siber. Mengamati setiap "bundel" informasi yang merupakan bagian dari pertukaran sistem usaha secara praktis tidak mungkin bagi pemeriksa manusia untuk menyaring secara tepat.

Pemrograman berbasis AI secara potensial dapat menggunakan berbagai prosedur, misalnya, analisis terukur, pencocokan kata kunci, dan identifikasi kekhasan untuk memutuskan apakah bundel informasi tertentu cukup unik dari tolok ukur paket informasi yang digunakan dalam kumpulan data persiapan. Semua ini tampaknya menunjukkan bahwa penalaran digital saat ini mulai dilihat sebagai instrumen yang efektif untuk meningkatkan pilihan nyata terhadap penipu dan peretas.

Kecerdasan Buatan untuk ID Bahaya Sistem

Keamanan sistem merupakan hal mendasar bagi sebagian besar organisasi, dan bagian tersulit dalam membangun bentuk keamanan siber sistem yang hebat adalah melihat semua komponen berbeda yang terkait dengan geologi sistem. Bagi spesialis keamanan siber manusia, ini berarti kerja keras dalam mengikuti semua korespondensi yang terjadi di seluruh rangkaian proyek. Berurusan dengan keamanan sistem proyek ini mencakup membedakan tuntutan asosiasi mana yang nyata dan mana yang berusaha melakukan perilaku asosiasi yang tidak biasa, misalnya, mengirim dan menerima sejumlah besar informasi atau memiliki proyek tidak teratur yang mengikuti asosiasi dengan rangkaian proyek.

Ujian bagi spesialis keamanan siber terletak pada pengenalan bagian mana dari suatu aplikasi, terlepas dari apakah di web, tahap seluler, atau aplikasi yang sedang dikembangkan atau diuji, yang mungkin jahat. Mengenali aplikasi jahat di antara banyak proyek serupa dalam rangkaian proyek berskala besar membutuhkan waktu yang sangat lama dan spesialis manusia umumnya tidak akurat. Pemrograman keamanan kerangka kerja berbasis penalaran buatan dapat menyaring semua lalu lintas kerangka kerja yang mendekat dan dinamis untuk mengenali model yang meragukan atau tidak biasa dalam data kemacetan waktu sibuk. Data

yang disinggung di sini umumnya terlalu banyak bagi otoritas keamanan digital manusia untuk secara tegas mengatur situasi risiko.

Pemantauan Email dengan Kecerdasan Buatan

Perusahaan ventura memahami pentingnya pemantauan pertukaran email untuk mencegah upaya peretasan keamanan siber, misalnya, phishing. Pemrograman pemantauan berbasis AI saat ini digunakan untuk membantu meningkatkan ketepatan penemuan dan kecepatan mengenali ancaman siber. Beberapa inovasi kecerdasan buatan yang berbeda digunakan untuk kasus penggunaan ini. Misalnya, beberapa produk menggunakan penglihatan komputer untuk "melihat" pesan guna memeriksa apakah ada fitur dalam email yang mungkin merupakan ciri bahaya, misalnya, gambar dengan ukuran tertentu. Dalam kasus lain, penanganan bahasa normal digunakan untuk membaca konten dalam pesan yang masuk dan keluar dari organisasi dan mengenali ekspresi atau contoh dalam teks yang terkait dengan upaya phishing. Menggunakan pemrograman lokasi inkonsistensi dapat membantu mengenali apakah pengirim, penerima, badan, atau tautan email merupakan bahaya.

Kasus penggunaan ini juga menampilkan kualitas kecerdasan berbasis komputer dengan analisis informasi skala besar. Tidak sulit bagi pekerja manusia untuk membaca email dan mengidentifikasi hal-hal yang mencurigakan; namun, melakukan hal tersebut untuk banyak pesan yang dikirim dan diterima dalam jaringan besar setiap hari pada dasarnya mustahil. Pemrograman kecerdasan berbasis komputer dapat membaca semua pesan yang masuk dan ramah dan melaporkan kejadian yang paling mungkin dari bahaya keamanan siber kepada staf keamanan. Misalnya, kasus untuk memberikan pemeriksaan email pemrograman kecerdasan buatan yang dapat memungkinkan perusahaan keuangan untuk mencegah pesan yang menyesatkan, mencegah penetrasi informasi dan serangan phishing. Produk organisasi kemungkinan menggunakan persiapan bahasa umum dan identifikasi kelainan dalam berbagai langkah untuk membedakan pesan mana yang mungkin merupakan bahaya keamanan siber.

Kecerdasan Simulasi untuk Memerangi Bahaya Kecerdasan Buatan

Organisasi perlu mengembangkan kecepatan dalam mengidentifikasi ancaman siber karena programmer saat ini menggunakan kecerdasan buatan untuk kemungkinan menciptakan resolusi bagian dalam klasifikasi profesional yang besar. Dengan cara ini, menugaskan pengembangan perangkat lunak kecerdasan buatan untuk membuat persiapan bagi upaya peretasan yang diperbesar dengan kecerdasan buatan dapat dikaitkan dengan bagian mendasar dari pertunjukan petugas keamanan digital di kemudian hari.

Dalam beberapa tahun terakhir, organisasi di seluruh dunia telah menyerah pada ancaman dunia maya dan serangan ransomware, misalnya, Winery dan notpetya. Kategori serangan ini menyebar dengan cepat dan memengaruhi banyak PC. Secara keseluruhan, penjahat yang bertanggung jawab atas jenis serangan ini dapat menerapkan revolusi kecerdasan berbasis komputer di kemudian hari. Sedikit kelonggaran yang dapat diberikan kecerdasan buatan kepada para programmer ini seperti yang disarankan kecerdasan buatan di perusahaan: fleksibilitas cepat. Serangan Cyber Security Merchant Gathering mengklaim pengembangan perangkat lunak retret mereka, Bird of prey Stage, menggunakan kecerdasan buatan untuk mempersiapkan ancaman ransomware tersebut. Penciptaan tersebut diduga

menggunakan penemuan inkonsistensi untuk retret titik akhir dalam klasifikasi komersial yang sangat besar.

Nasib Kecerdasan Berbasis Komputer dalam Keamanan Siber

Penggunaan pengetahuan buatan dalam struktur keamanan digital dapat disebut sejak saat ini. Asosiasi perlu memastikan bahwa struktur mereka disiapkan dengan komitmen dari para profesional keamanan digital yang akan meningkatkan kemampuan mengenali penyergapan canggih yang sebenarnya dengan presisi yang jauh lebih tinggi daripada sistem keamanan digital biasa. Asosiasi perlu memahami bahwa sistem ini berada dalam kelas yang sama dengan data yang ditanganinya. Sistem kecerdasan buatan biasanya secara luas ditingkatkan menjadi struktur "masuk sampah, keluar sampah", dan pendekatan berbasis data untuk mengelola petualangan wawasan berbasis PC adalah hal utama untuk kemajuan yang berkelanjutan.

Satu-satunya ujian bagi asosiasi yang pada dasarnya menggunakan metode pengungkapan keamanan digital berbasis kesadaran buatan adalah mengurangi jumlah pengungkapan positif palsu. Ini mungkin menjadi lebih mudah dilakukan karena item tersebut mengenali apa yang telah ditandai sebagai laporan positif palsu. Pada saat standar langsung telah dibuat, hitungan dapat secara autentik menandai penyimpangan yang signifikan sebagai kekhasan dan menyiapkan inspektur keamanan bahwa penilaian tambahan diperlukan. Pengajuan keamanan digital adalah salah satu permintaan AI yang paling terkenal saat ini. Hal ini sebagian besar disebabkan oleh cara aplikasi ini bergantung pada pengakuan kekhasan yang sangat cocok untuk model AI. Selain itu, sebagian besar asosiasi besar mungkin mulai sekarang memiliki pertemuan keamanan digital yang ada, rencana pengeluaran kemajuan hal, dan pendirian TI untuk mengelola banyak data.

3.8 CARA MENINGKATKAN KEAMANAN SIBER UNTUK KECERDASAN BUATAN

Ringkasan strategi ini menyelidiki isu-isu utama dalam upaya meningkatkan keamanan siber dan keamanan untuk kesadaran buatan serta pekerjaan bagi para pembuat kebijakan dalam membantu mengatasi kesulitan-kesulitan ini. Kongres baru saja menunjukkan antusiasmenya terhadap pemberlakuan keamanan siber dengan berfokus pada jenis inovasi tertentu, termasuk *Internet of Things* (IOT) dan kerangka kerja pemungutan suara. Karena kecerdasan simulasi berubah menjadi inovasi yang semakin signifikan dan umum digunakan di berbagai bidang, para pembuat kebijakan akan merasa semakin penting untuk memikirkan titik temu keamanan siber dengan berbasis AI.

Saya menggambarkan sebagian isu yang muncul di sekitar sana, termasuk pertukaran kerangka kerja dinamis kecerdasan buatan untuk tujuan-tujuan yang berbahaya, potensi musuh untuk mengakses informasi atau model pelatihan kecerdasan buatan pribadi, dan usulan strategi yang direncanakan untuk menangani kekhawatiran ini. Mengamankan Sistem Pengambilan Keputusan Kecerdasan Buatan: Salah satu ancaman keamanan kritis terhadap sistem penalaran buatan adalah potensi musuh untuk menangani teknik dinamis mereka sehingga mereka tidak membuat pilihan dengan cara yang diharapkan atau dibutuhkan oleh para insinyur mereka. Salah satu cara untuk mencapainya adalah dengan meminta musuh

untuk benar-benar menerima tanggung jawab atas struktur pengetahuan buatan manusia dengan tujuan agar mereka dapat memilih hasil yang dihasilkan sistem dan keputusan apa yang diambilnya. Di sisi lain, penyerang dapat berupaya memengaruhi keputusan tersebut dengan lebih langsung dan tidak langsung dengan meneruskan sumber data yang beracun atau menyiapkan data ke model pengetahuan berbasis komputer.

Misalnya, musuh yang ingin menangani kendaraan yang dapat mengatur dirinya sendiri dengan tujuan agar kendaraan tersebut tidak diragukan lagi akan mengalami kemunduran dapat menyalahgunakan kekurangan pada item kendaraan untuk memilih keputusan mengemudi sendiri. Namun, secara tidak langsung mengakses dan menyalahgunakan item yang mengoperasikan kendaraan dapat terbukti menjengkelkan, jadi dengan mempertimbangkan semuanya, musuh dapat berupaya membuat kendaraan tersebut mengabaikan rambu berhenti dengan menghancurkannya di area tersebut dengan cat semprot. Dengan demikian, penghitungan penglihatan PC tidak akan memiliki alternatif untuk mengingatnya sebagai tanda berhenti. Strategi yang digunakan musuh untuk membuat struktur pengetahuan berbasis PC melakukan kesalahan dengan mengendalikan sumber data ini disebut serangan udara. Para peneliti telah menemukan bahwa perubahan kecil pada gambar mutakhir yang tidak jelas bagi mata normal dapat cukup untuk membuat perhitungan wawasan yang direkayasa benar-benar salah mengklasifikasikan foto-foto tersebut.

Cara elektif untuk menangani pengendalian sumber informasi adalah perusakan informasi, yang terjadi saat musuh melatih model kecerdasan buatan pada informasi yang salah diberi label. Gambar rambu berhenti yang ditandai sebagai sesuatu yang berbeda sehingga kalkulasi tidak akan mengenali rambu berhenti saat melihatnya di luar adalah contohnya. Perusakan model ini kemudian dapat menyebabkan kalkulasi kecerdasan buatan melakukan kesalahan dan kesalahan klasifikasi di kemudian hari, terlepas dari apakah musuh tidak secara sah mengendalikan sumber informasi yang diterimanya. Bahkan hanya dengan menyiapkan model kecerdasan buatan secara khusus pada sebagian kecil informasi yang diberi nama secara akurat mungkin cukup memadai untuk menawar model dengan tujuan agar model tersebut memutuskan pilihan yang salah atau mengejutkan.

Saat ini, orang-orang hidup dalam masyarakat maju di mana data atau informasi yang keras dan cepat disimpan dalam bentuk elektronik/online. Informasi tersebut mungkin terkait dengan kehidupan pribadi, perdagangan terkait uang, kemajuan hukum, atau informasi lain apa pun yang sifatnya penting. Tentunya, kumpulan informasi telah disajikan di area korespondensi individu tanpa memahami risiko keamanannya. Ini adalah ciri khas para perusuh komputer karena informasinya dapat diakses secara terbuka. Keamanan digital bukan hanya masalah yang berkaitan dengan seseorang. Ini juga berlaku untuk sebuah organisasi. Setiap saat seseorang harus dapat menjamin data atau informasi tentang tujuan koneksi relasional, dan informasi yang terkait dengan transaksi perbankan harus memiliki upaya keamanan yang cukup.

Keamanan siber bukan hanya divisi inovasi data atau masalah atau kewajiban yang menyangkut individu di kantor yang sama. Ini adalah aktivitas setiap pekerja dan bahkan klien asosiasi. Menurut perayap web Google, orang-orang online seperti jarum jam setiap menit

setiap hari. Jadi, bagaimana kita dapat melindunginya? Peraturan Perlindungan Data Umum (GDPR), sebuah peraturan dalam hukum Uni Eropa, memberi individu lebih banyak kendali atas data pribadi mereka.

Namun, ada asosiasi yang telah mengalami serangan siber, akan mengalaminya dan mungkin telah mengalaminya, tetapi tidak memiliki ide sedikit pun tentang apa yang harus dilakukan tentangnya. Untuk menemukan jawaban yang lebih baik tentang ini, kita memerlukan strategi intelijen berbasis komputer. Memahami hubungan antara kecerdasan atau sains berbasis komputer yang dapat menyalin individu dan keamanan siber yang merupakan persyaratan mendasar untuk segala hal adalah cara untuk mencapai keberhasilan dalam bisnis saat ini. Ada sistem yang tersedia untuk menjamin informasi di internet, seperti keamanan kata rahasia, konfirmasi data, pemindai malware, firewall, pemrograman antivirus, dll. Dengan menyadari etika tingkat lanjut yang autentik, sejumlah besar penyergapan terkomputerisasi dapat dilawan. Namun, pelanggaran atau serangan terkomputerisasi terus berkembang biak di berbagai jalur dari waktu ke waktu. Tidak ada tindakan yang sempurna atau jawaban menyeluruh untuk perilaku buruk/penyergapan terkomputerisasi, namun perangkat terbaru mampu membatasinya untuk memastikan keamanan di dunia yang maju.

BAB 4
DETEKSI BOTNET MENGGUNAKAN KECERDASAN BUATAN

Selama 10 - 15 tahun terakhir, botnet menjadi perhatian utama dalam dunia keamanan siber. Botnet adalah jaringan perangkat yang terinfeksi malware jenis *bot*, dikendalikan oleh *botmaster* untuk melakukan aktivitas berbahaya seperti pencurian data atau serangan DDoS. Penelitian menunjukkan bahwa metode deteksi tradisional, seperti sistem deteksi intrusi berbasis tanda tangan, kurang efektif menghadapi botnet modern yang semakin canggih dan tersembunyi. Oleh karena itu, pendekatan berbasis *machine learning* mulai banyak digunakan karena mampu mengenali pola perilaku mencurigakan tanpa bergantung pada tanda tangan yang telah diketahui. Bab ini mengulas literatur terkait dan mengusulkan metode baru berbasis algoritma optik untuk mendeteksi dan memahami jenis botnet secara lebih akurat. Inovasi ini mencakup pelacakan IP botmaster, analisis akses layanan oleh botnet, serta penggabungan fitur perilaku terbaru. Pendekatan ini diharapkan mampu meningkatkan akurasi dan efektivitas deteksi botnet dalam lanskap ancaman digital yang terus berkembang.

4.1 PENGENALAN BOTNET

Istilah botnet berasal dari ungkapan otomatis dan jaringan. Bot, yang kadang-kadang dikenal sebagai zombi, adalah perangkat karakter yang terhubung ke pengaturan konvensi web (IP), biasanya web. Umumnya, ini berarti mendaftarkan perangkat komputer, komputer, printer, sakelar rumah tangga, dll. yang wajib berubah menjadi bot. Namun, saat ini, seiring dengan kemajuan *Internet of Things* (IoT), perangkat keluarga kita semakin banyak yang secara rutin dikaitkan dengan web. Yang berarti daftar perangkat botnet yang semakin canggih pada dasarnya telah digandakan. Yang sekarang dilindungi adalah kamera web. Setelah perangkat terinfeksi malware botnet, perangkat tersebut dapat digunakan melalui jaringan lokalnya untuk mengarahkan sejumlah besar permainan yang tidak sah dan berbahaya.

Setelah perangkat diinstal dengan "perangkat lunak bot" melalui infeksi malware, "bot herder" dapat membuat bot melakukan apa saja dengan mengeluarkan perintah melalui server perintah dan kontrol (C&C atau C2). Botnet biasanya terdiri dari ratusan atau bahkan jutaan perangkat, termasuk PC, Mac, server Linux, router rumah, telepon pintar, dll. Sumber daya gabungan dari perangkat yang dikendalikan ini dapat digunakan untuk meluncurkan serangan yang merusak atau canggih seperti mengirim miliaran email spam, DDoS bandwidth besar, dan penipuan keuangan yang ditargetkan. Botnet secara tradisional menggunakan protokol HTTP dan IBN untuk berkomunikasi dengan klien botnet yang terinfeksi.

Gambar 4.1 Pengenalan Botnet.

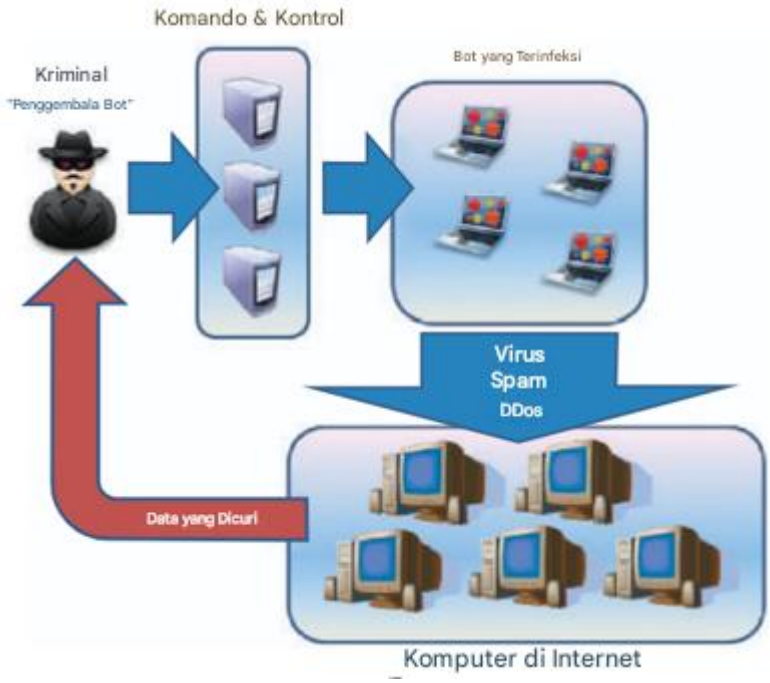
Untuk memblokir komunikasi ini, layanan keamanan jaringan dapat mengontrol akses ke layanan dan port ini. Misalnya, Firebox dapat menggunakan kategori Perintah dan Kontrol WebBlocker dan Aktivitas Botnet untuk memblokir komunikasi dari klien botnet yang terinfeksi di jaringan Anda ke situs botnet melalui HTTP. Gambar 4.1 di atas menunjukkan ikhtisar Botnet tempat semua komponen yang diperlukan telah diperkenalkan. Penggembala botnet adalah penghibur yang mengawasi bot dari jarak jauh. Mereka mengatur dan menetapkan pekerja perintah dan kontrol (C&C), yang mengisi antarmuka ke bot. Penggembala botnet adalah malware yang mengawasi bot dari jarak jauh. Mereka mengasosiasikan dan menetapkan karyawan perintah dan kontrol (C&C), yang mengisi antarmuka ke bot. Misalnya, saluran IBN sering digunakan untuk alasan itu. Setelah komunikasi disiapkan, host yang dinegosiasikan sering kali dipersiapkan lebih lanjut dan diberikan instruksi yang diperbarui. Mereka sekarang telah muncul sebagai kelompok host yang dipersiapkan di bawah manajemen terpusat.

4.2 DETEKSI BOTNET

Seiring munculnya botnet sebagai ancaman yang lebih besar, para peneliti dan pakar keamanan mengembangkan teknik dan strategi khusus untuk mengatasi masalah tersebut. Metode deteksi mendefinisikan cara kerja jawaban, bersama dengan deteksi berdasarkan perilaku atau tanda tangan yang ditunjukkan pada Gambar 4.2. Strategi luar biasa terutama didasarkan pada strategi luar biasa.

Berbagai teknik telah dirancang untuk deteksi Botnet dari waktu ke waktu. Tiga teknik utama deteksi botnet dijelaskan di bawah ini:

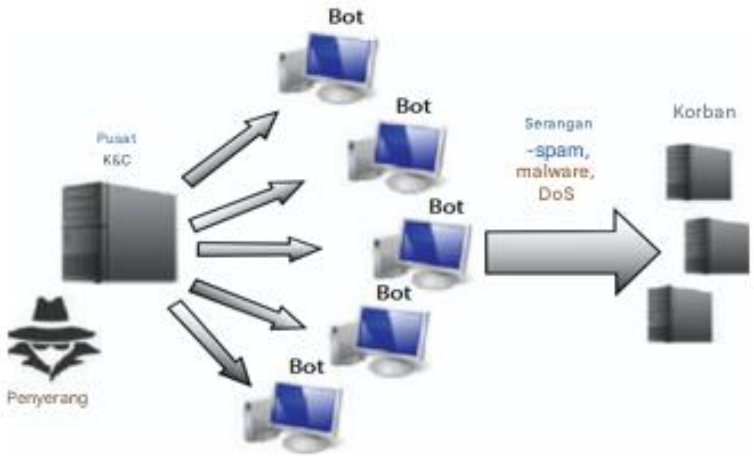
a) Deteksi Berpusat pada Host
b) Deteksi Berbasis Honey Nets
c) Deteksi Berbasis Jaringan



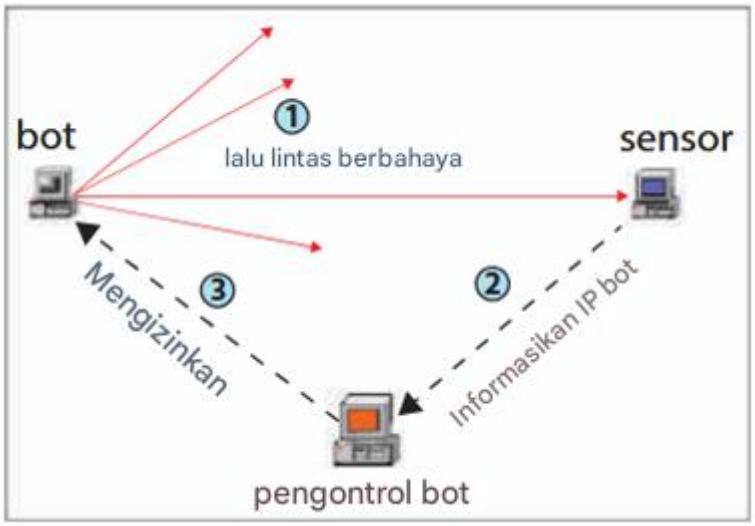
Gambar 4.2 Faktor Botnet.

Deteksi Berpusat pada Host (HCD)

Dalam metode identifikasi berbasis HCD, pemeriksaan reaksi mesin dilakukan berdasarkan ketentuan tertentu. Perilaku keseluruhan mesin diawasi dan dilakukan upaya untuk menemukan segala jenis variasi dari norma. Ini termasuk kerangka kerja yang membutuhkan waktu terlalu lama untuk berpikir tentang menanggapi bahkan aktivitas kecil, membutuhkan waktu terlalu lama untuk mempertimbangkan penyelesaian suksesi panggilan, bagian yang meragukan di perpustakaan, perubahan yang tidak biasa dalam kerangka kerja rekaman, antivirus tidak bereaksi atau membunuh sendiri, perubahan dalam asosiasi pengaturan dan sebagainya dapat menunjukkan keberadaan bot. Strategi penemuan yang berpusat pada host tidak dianggap sebagai teknik yang sangat efektif karena strategi tersebut cocok untuk satu mesin dan dapat berubah dari satu mesin ke mesin lainnya yang ditunjukkan pada Gambar 4.3.



Gambar 4.3 Sistem deteksi yang berpusat pada host.



Gambar 4.4 Deteksi botnet berbasis honey nets.

Deteksi Berbasis Honey Nets (HNBD)

Honey nets (kadang-kadang juga disebut honeypots) sebagian besar digunakan untuk merenungkan dan memahami sorotan dan metode botnet; namun, honeynets umumnya tidak berharga dalam mengidentifikasi kontaminasi bot. Honey nets biasanya digunakan untuk menemukan target ahli atau penyerang bot. Metode ini berharga untuk mengidentifikasi bot yang dikenal. Bot yang tidak dikenal dan bahkan bot yang dikenal dengan sedikit perubahan pada paralel bot tidak diidentifikasi oleh strategi ini yang ditunjukkan pada Gambar 4.4.

Deteksi Berbasis Jaringan (NBD)

Metode lokasi gabungan sistem didasarkan pada pemeriksaan dan pemecahan lalu lintas sistem yang tidak aktif. Metodologi ini sangat berguna dalam mengenali keberadaan botnet dalam sistem. Dalam metodologi ini, informasi sistem diperiksa secara konsisten, komunikasi berbasis organisasi dipantau. Setiap tindak lanjut yang tidak lazim dapat menunjukkan adanya tindakan balas dendam. Sekarang para pembual itu cerdas dan menerapkan sejumlah besar metode pengacauan kode. Meskipun kode jahat itu diacak dan dielakkan oleh perangkat lunak pendeteksi malware, paket-paket itu masih ada dalam sistem, dan dapat juga diikuti dengan menerapkan metode yang berbeda.

4.3 ARSITEKTUR BOTNET

Bagian bab ini menunjukkan beberapa fitur utama botnet. Botnet adalah komunitas yang mencakup mesin-mesin di bawah kendali langsung seorang operator tertentu yang dikenal sebagai botmaster. Botnet dapat dianggap sebagai penggabungan berbagai bahaya menjadi satu. Botnet umumnya berisi server botnet yang membangun ratusan ribu botnet dengan jaringan. Ratusan ribu botnet ini dikenal sebagai botnet kecil, dan botnet yang berisi jutaan klien dikenal sebagai botnet besar.

Istilah botnet dipinjam dari "*Robotic Community*". Istilah ini mencerminkan kenyataan bahwa klien bot akan berperilaku seperti robot, dan server, yaitu botmaster, digunakan untuk mengirimkan instruksi dan memenuhi tujuan meluncurkan serangan ke satu lokasi pusat. Saat

ini, seorang botmaster tunggal menangani banyak server bot dengan menyiapkan berbagai segmen. Botmaster biasanya menempatkan percakapan dengan bot menggunakan IBN (*Internet Broadcast Negotiation*) pada server perintah dan kontrol yang jauh. Ini terdiri dari lima level utama yang ditampilkan di dalam percakapan mulai dari menghubungkan bot baru hingga meluncurkan serangan, seperti yang dijelaskan di bawah ini:

- Perangkat baru yang tersedia diserang dan didamaikan dengan menyalin kode berbahaya ke dalamnya. Ketika kode berbahaya dieksekusi, mesin mencari budak C&C, yang terhubung ke server dan menjadi bagian dari 72 botnet. Prosedur pengumpulan menginformasikan keberadaan botnet dan kemudian botmaster siap menerima perintah.
- Setelah itu, botmaster menerima perintah dari botmaster untuk menjalankan beberapa misi jahat.
- Perintah yang diterima melalui mesin dari botmaster kemudian dicapai dengan menggunakan konsumen bot.
- Serangan diperkenalkan sesuai dengan perintah yang diberikan.
- Pelindung bot menanggapi botmaster untuk memberitahunya tentang pemenuhan serangan.

Model Federal

Dalam model federal, ada satu pekerja paling signifikan atau komponen penting yang bertanggung jawab untuk menyiapkan perdagangan verbal antara pelanggan yang terganggu dan bot. Pemanfaatan saluran ini, perdagangan pesan dan arahan dilakukan. Pekerja utama dikenal sebagai pekerja perintah dan kontrol (C&C). Banyak botnet yang dapat diakses di sepanjang tepi sabot, bot lalu, robot, dll., menggunakan C&C untuk penjelasan diskusi. PC atau pekerja dasar sering kali merupakan gawai PC yang luar biasa karena harus menangani seluruh botnet yang ukurannya juga dapat berfluktuasi dari beberapa paket hingga beberapa tumpukan ratusan. Ia harus memiliki kapasitas transfer data yang tinggi mengingat pada satu titik ia mungkin perlu melayani banyak bot. Meskipun pekerja dasar adalah yang luar biasa, ia dianggap sebagai faktor yang cenderung dari versi ini. Ada beberapa konvensi yang dapat digunakan secara teratur oleh C&C untuk melakukan korespondensi dan itu adalah HTTP (konvensi transfer konten hiperliterasi) dan IBN.

Protokol Berbasis IBN

IBN adalah konvensi untuk penyampaian teks web secara real-time atau konferensi terkoordinasi yang bergantung pada TCP yang juga dapat memanfaatkan lapisan lampiran yang longgar. IBN memberikan berbagai kemampuan yang bermanfaat. Protokol ini memungkinkan pemindahan berkas di antara pelanggan dan paket yang berjalan di struktur. Botnet berbasis IBN menggunakan perintah terpusat dan bentuk manipulasi di mana mesin yang terinfeksi mencoba dan membuat koneksi ke server IBN dan menjadi bagian dari saluran yang sama. Dalam botnet ini, server C&C bekerja pada perusahaan IBN. Protokol IBN terutama didasarkan pada model pelanggan-server. Protokol ini memberikan fleksibilitas dalam percakapan dan cukup mudah untuk disiapkan. Protokol ini adalah salah satu protokol paling populer untuk menempatkan pertukaran verbal di antara botnet.

Botnet Berbasis HTTP

IBN dapat dianggap sebagai konvensi awal untuk botnet. IBN memperoleh keunggulan besar dan sebagian besar botnet bekerja pada IBN, sehingga banyak analis mulai menghabiskan banyak waktu dalam laporan berbasis IBN. IBN juga memiliki beberapa kekurangan. Karena IBN terdiri dari fakta tentang jangkauan luas port sebelum melakukan serangan, serangan dapat dideteksi tanpa kesulitan. Jadi para peretas beralih ke protokol HTTP. Protokol ini umumnya digunakan dalam kategori komunitas apa pun. Ia menawarkan banyak keuntungan. Ia memiliki fungsi untuk menyembunyikan pengunjung situs web botnet jahat di pengunjung situs web biasa yang tidak dapat dideteksi melalui firewall. Botnet berbasis HTTP sepenuhnya mudah dibentuk dan dijalankan. Ada beberapa botnet yang menggunakan protokol HTTP dan mereka adalah restock, clickbot, Zeus, ada beberapa gaya botnet berbasis HTTP yang benar-benar nyata:

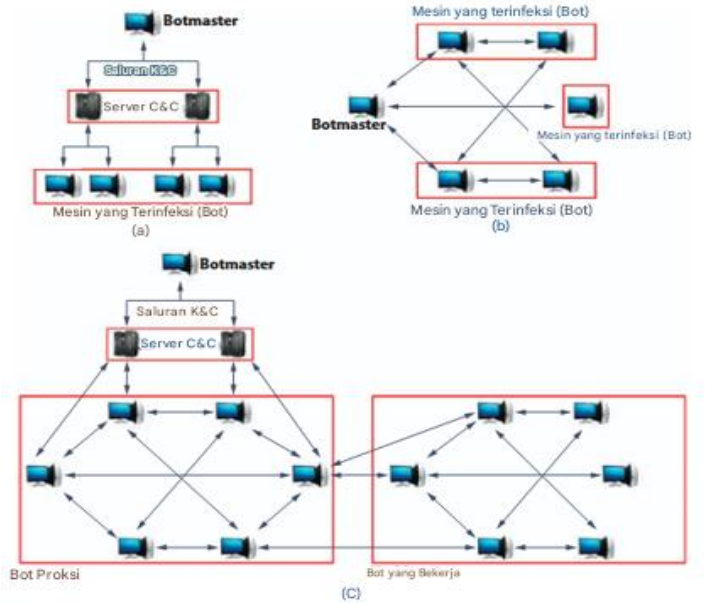
1. Botnet HTTP yang berpusat pada resonansi
2. Botnet HTTP yang berpusat pada pengetahuan

Model Terdesentralisasi

Sebuah model dikatakan terdesentralisasi ketika tidak ada perintah dan manipulasi yang signifikan. Kebiasaan yang digunakan oleh botnet terpusat adalah IBN dan HTTP meskipun botnet ini bekerja dengan beberapa jenis protokol. *Peer to peer* (P2P) adalah contoh dari versi desentralisasi. Komunitas P2P dari mesin penyelesaian cukup kuat untuk diketahui dan dihancurkan. Mesin ini biasanya menggunakan jaringan berbagi dokumen. Dalam mode desentralisasi, komputer bebas memilih bot mana pun untuk mendistribusikan perintah di dalam botnet. Semua bot dapat bertindak sebagai pelanggan lebih lanjut ke server. Bentuk botnet ini tidak dapat dihancurkan hanya dengan menyerang satu hal karena tidak ada server penting untuk memanipulasi seluruh botnet. Jika satu bot diserang dan dihancurkan, maka bot lain dari botnet akan tetap beroperasi. Botnet P2P lebih dinamis dan kuat daripada yang tersentralisasi. Setiap bot menjaga beberapa kolaborasi dengan bot lain dari botnet. Botnet P2P cukup sulit untuk dipantau, dihancurkan, dan diretas.

Model Silang

Arsitektur hibrida merupakan campuran dari struktur terpusat dan terdesentralisasi. Dinyatakan bahwa dengan arsitektur hibrida terdapat dua jenis bot. Bot klien dan bot pelayan. Pelacakan dan pendeteksian botnet yang memiliki struktur hibrida lebih sulit dibandingkan dengan arsitektur terpusat dan terdesentralisasi yang ditunjukkan pada Gambar 4.5 (Pramod Singh Rathore et al., 2017).



Gambar 4.5 Arsitektur botnet (a) model federal (b) model yang didelegasikan (c) arsitektur silang.

Deteksi Botnet

Dalam teknik deteksi, solusi akan beroperasi sesuai dengan sifatnya. Strategi deteksi berbasis pembelajaran mesin mampu menggunakan kedua teknik tersebut. Strategi unik yang diterapkan dalam deteksi bot meliputi anomali dan sistem nama area.

Perspektif Deteksi Botnet

Strategi berbasis tanda mencari informasi yang berbeda pada bot atau atribut yang diidentifikasi dengan bot, seperti Lalu Lintas, mungkin juga tampak seperti itu. Pendekatan berbasis tanda tangan mencari informasi terbaik tentang bot atau yang terkait dengan ciri bot, seperti Pengunjung Situs, juga dapat tampak seperti itu. Metodologi ini digunakan untuk perkembangan tertentu termasuk aturan atau administrasi tertentu. Metodologi semacam ini beralih ke yang spesifik dan khusus, sesuatu di luar ekstensi yang diinginkan akan bergerak tak terpetakan. Teknik ini mungkin terlalu kuat untuk melawan botnet yang teridentifikasi, tetapi sekarang ini tidak berguna untuk bot anonim dan lebih rentan terhadap strategi elusi.

Pendekatan berbasis deteksi sebagian besar didasarkan pada perilaku bot dan terdiri dari mendeskripsikan versi untuk cara botnet biasanya berfungsi. Kesederhanaan pendekatan ini memungkinkan untuk menangkap bot baru atau yang tidak mencolok, tetapi, sangat elegan dan tuduhan palsu yang beredar mungkin menjadi tidak wajar. Dalam teknik perilaku, peneliti membuat asumsi berdasarkan pengamatan tentang perilaku inti botnet. Untuk deteksi botnet, asumsi nomor satu di awal teknik adalah bahwa bot bekerja secara kooperatif, terlibat dalam beberapa bentuk minat kelompok di berbagai tingkat siklus keberadaan botnet. Di mana statistik khusus bot tertentu mendorong pendekatan berbasis tanda tangan benar-benar merupakan definisi yang lancar tentang perilaku bot yang menjadi inti dari strategi perilaku.

Teknik Deteksi (Pengungkapan)

Strategi ABD (*Anomaly-Based Detection*) bertujuan untuk menemukan bot yang bergantung pada olahraga web yang luar biasa, yang mencakup transportasi yang sangat

berlebihan, potensi besar, dan aksi olahraga pelabuhan yang tidak biasa. Teknik ABD mengambil strategi perilaku untuk identifikasi bot, oleh karena itu, ia dapat memilih permainan yang menakjubkan atau melakukan untuk bot yang tidak dikenal. Teknik berdasarkan DNF melakukan fakta yang dibentuk dengan sumber daya botnet. Saluran percakapan C&C unik untuk malware bot; bot berinteraksi dengan server C&C melalui saluran tersebut. Untuk mendapatkan akses ke hub tersebut, bot menjalankan kueri DNF.

Tujuan dari strategi berbasis DNF adalah untuk membuat pengunjung situs DNF yang tidak biasa melihat bot. AI yang pada dasarnya tunduk pada kerangka kerja pengenalan telah dipandang sebagai yang terbaik dalam membedakan botnet. Efisiensinya terletak pada potensinya untuk mengetahui lalu lintas reguler internal pengunjung situs yang terkait dengan bot. Itu adalah tugas untuk strategi yang berbeda karena bot menggunakan konvensi biasa untuk membuat laporan C&C. Namun, deteksi pembelajaran mesin memerlukan jumlah kasus pembelajaran yang sesuai dan kemampuan yang dijelaskan dengan tepat agar dapat beroperasi. Sebagai inti dari bab ini, pembelajaran perangkat akan disebutkan secara lebih rinci di bagian berikut.

Wilayah Penelusuran

Bot pembeda yang bergantung pada pergerakan usaha asosiasi menerima permainan yang disusun melalui bot di dalam botnet yang setara. Melalui gaya lalu lintas situs web yang sebanding, tujuannya adalah untuk memahami semua bot dalam komunitas tersebut sebagian besar berdasarkan pergerakan kolektif mereka sebagai pengganti operasi karakter mereka. Dibandingkan dengan deteksi berbasis institusi, kawanan manusia diberi label berdasarkan aktivitas dan sifat karakter mereka terlepas dari minat institusi tempat mereka mungkin menjadi bagiannya.

4.4 PEMBELAJARAN MESIN

Pembelajaran mesin adalah aplikasi AI yang bertujuan untuk meningkatkan sistem yang mampu menganalisis dari pengalaman masa lalu. Dalam pembelajaran perangkat, fakta masa lalu diberikan sebagai masukan ke algoritma pembelajaran perangkat untuk mengumpulkan gaya yang mungkin ada dengan tujuan untuk membangun model yang menunjukkan catatan. Pada titik fokus ML terdapat pemikiran terukur dan komputasional yang diperoleh dari aturan yang ada dalam banyak tatanan yang meliputi kesadaran buatan manusia, teori, pemikiran pengukuran, sains, keterampilan inovatif psikologis, sifat komputasional yang multifaset, dan undang-undang kontrol. Tujuan ML adalah untuk membentuk versi berdasarkan statistik yang diberikan. Versi ini menggambarkan gaya yang ada dalam catatan yang seharusnya sebagai cara untuk membuat pilihan yang berpengetahuan berdasarkan fakta baru (yang tidak terlihat).

Karakteristik Pembelajaran Mesin

Teknik pembelajaran mesin digunakan untuk melacak botnet. Fitur-fitur tersebut menentukan bentuk versi yang dibuat. Fungsi mampu menunjukkan perilaku atau tujuan dari ciri-ciri tertentu. Metode ML yang dipilih akan memengaruhi perilaku model; satu pendekatan juga dapat lebih jauh menciptakan versi yang tantangan utamanya adalah bagaimana bot

tertentu berinteraksi satu sama lain sementara setiap model lainnya memikirkan sendiri bagaimana karakteristik bot individu. Untuk mengekstrak pembagian variabel yang optimal melalui semua variabel yang memungkinkan yang menunjukkan informasi yang paling akurat, metode yang disebut pilihan kapabilitas digunakan. Di lokasi botnet, motivasi di balik metode keputusan aktivitas adalah untuk memilih subset kapasitas sementara dalam transit untuk menggambarkan secara maksimal bot atau bot terbaik yang sedang dipantulkan.

Kompetensi yang diputuskan akan sepenuhnya didasarkan pada bentuk informasi yang digunakan. Berbagai jenis pencarian kueri dapat berupa fungsi dari statistik DNF, pengiriman dan lokasi ip untuk catatan aliran internet, checksum adalah fungsi dari statistik top paket. Untuk deteksi berbasis ML, sebagian besar peneliti memilih aliran internet. Contoh kemampuan tingkat waft adalah: waktu aliran, byte umum per paket per drift, yang menunjukkan hubungan antara pengguna dan server.

Kemampuan yang dipilih akan memberikan pendekatan tertentu. Kemampuan tingkat waft akan membantu teknik perilaku, kompetensi tingkat paket yang menangkap tren tertentu akan membantu teknik berbasis tanda tangan. Hipotesis dasar untuk penemuan botnet berbasis AI pada dasarnya adalah, bot membuat pola tertentu yang tidak terlihat dalam pengunjung situs web atau membeli aktivitas aplikasi. Oleh karena itu, dengan menerapkan beberapa bentuk teknik ML, seseorang mungkin dapat menemukan pola tersebut untuk secara efektif menemukan aktivitas jahat.

Pendekatan Pembelajaran Mesin untuk Mendeteksi Botnet

Awalnya, strategi maksimum yang dipasang untuk menggagalkan botnet bersifat reaktif, sehingga mengurangi efektivitasnya secara signifikan. Penelitian hebat telah dilakukan baru-baru ini untuk strategi yang lebih proaktif yang ditujukan untuk mengatasi botnet. Metode proaktif menyelidiki dinamika botnet dalam upaya untuk memahami siklus, fungsi, struktur, desain, dan pola serangannya serta metode otomatis dan waktu nyata untuk mengidentifikasi dan mendeteksi botnet. Hipotesis mendasar untuk studi sistem yang terutama bergantung pada deteksi botnet adalah bahwa bot membuat desain tertentu yang tersembunyi dalam pergerakan komunitas atau peristiwa perangkat patron. Menegakkan algoritme studi perangkat mungkin ingin membantu menemukan pola tersembunyi tersebut untuk secara efektif menyerang aktivitas jahat.

Banyak proses telah menggunakan serangkaian MLA yang digunakan dalam berbagai pengaturan. Metode ini menggunakan beberapa prinsip analisis lalu lintas yang difokuskan pada berbagai sifat aktivitas jaringan botnet. Selain itu, strategi deteksi saat ini dievaluasi menggunakan metodologi penilaian khusus dan unit fakta. Berbagai solusi deteksi yang beragam dan luar biasa memperkenalkan keinginan akan pendekatan yang komprehensif untuk meringkas dan mengevaluasi upaya klinis saat ini, dengan tujuan untuk menginformasikan tantangan keanggunan teknik deteksi ini dan menentukan kemungkinan untuk masa depan.

Beberapa penulis telah mencoba meringkas bidang keamanan botnet melalui serangkaian makalah survei. Secara paralel, beberapa penulis telah meringkas upaya klinis pada deteksi botnet dengan cara menawarkan taksonomi baru metode deteksi dan sejumlah

strategi yang paling menonjol. Para penulis telah menceritakan kapasitas taktik berbasis penguasaan gadget dalam menyediakan deteksi yang efisien dan kuat. Teknik Pembelajaran Mesin (ML) yang digunakan dalam deteksi botnet dapat dibahas di bawah dua klasifikasi utama pembelajaran mesin, khususnya strategi pembelajaran perangkat yang diawasi dan tidak diawasi.

4.5 DETEKSI BOTNET DENGAN PEMBELAJARAN MESIN

Fase ini mengevaluasi kontribusi setiap kompetensi dalam deteksi virus berdasarkan bagaimana mereka telah digunakan. Penilaian setiap teknik akan dipisahkan sebagai berikut: pertama, definisi pendekatan dan kedua, deskripsi singkat tentang bagaimana metode tersebut telah digunakan.

Pembelajaran Terbimbing (Diatur)

Dalam jenis pembelajaran ini, model dibangun pada rekaman pembelajaran berlabel. Tujuannya adalah membuat versi (tugas h) yang mewakili rekaman, yang didefinisikan oleh tugas (h) yang memetakan variabel x ke tujuan yang sesuai y, karakteristik ini kadang-kadang dilambangkan sebagai spekulasi $h(x)$. Ada perbandingan antara strategi pembelajaran terbimbing berdasarkan informasi berlabel.

Pemahaman masalah pembelajaran perangkat terbimbing dapat diklasifikasikan sebagai pembalikan atau jenis. Untuk masalah pembalikan, harga tujuan berlabel merupakan rentang nilai yang cukup besar. Untuk masalah ini, variabel input ditetapkan ke sekolah terutama berdasarkan gaya yang direpresentasikan dalam rekaman. Prosedur kelas terlibat oleh hubungan antara tag grup dan variabel input. Penelusuran botnet adalah contoh dari masalah tersebut, di mana kita mencoba memilih paket kelas bundel mana yang mungkin dibagikan, yaitu, botnet atau bukan pengunjung situs botnet.

Munculnya Pembelajaran Terbimbing

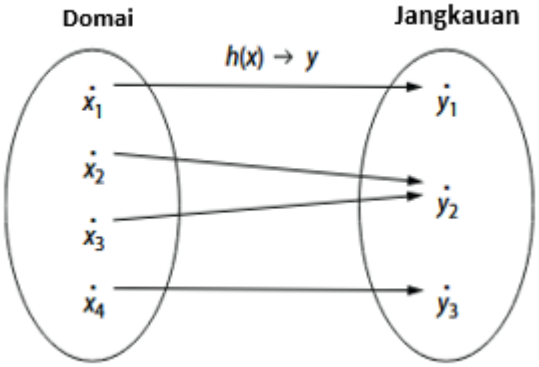
Pada tahun 2006, pendekatan berbasis deteksi bot-net berbasis jaringan diperkenalkan. Penulis memastikan penilaian terhadap tiga teknik pembelajaran gawai yang berbeda untuk mengetahui bot-net IRC. Deteksi dicapai secara bertahap; fase utama mengklasifikasikan pengunjung situs terutama berdasarkan lalu lintas IRC. Fase kedua mengklasifikasikan aliran obrolan IRC.

Bot IRC adalah pendekatan lain dari deteksi botnet. Metode ini dibagi menjadi lima tingkat. Pada tingkat pertama, aliran yang kemungkinan besar tidak memerlukan catatan C&C disaring sepenuhnya berdasarkan pemahaman bot IRC, gaya perilaku, dan sifat dalam aliran. Fase kedua menggunakan pembelajaran terbimbing untuk menyadari aliran lalu lintas yang mencurigakan. Asosiasi 0,33 derajat berjalan bergantung pada atribut pradefinisi yang sebanding. Kelompok tersebut kemudian dilampaui ke tahap keempat yang menggunakan evaluasi topologi untuk memilih aliran dengan regulator yang tidak dapat dibedakan. Aliran dengan regulator yang tidak dapat dibedakan kemudian dianalisis untuk memeriksa apakah aliran tersebut mungkin merupakan aspek botnet atau tidak.

Teknik ketiga diperkenalkan untuk mengetahui bot P2P. Tingkat pertama dari metode ini memunculkan ekstraksi fitur. Pada tingkat ini, fungsi unik yang dapat digunakan untuk

melambangkan bot P2P diambil dari proses pengunjung. Kompetensi aliran ini diserahkan ke tingkat kedua di mana serangkaian peraturan pembelajaran yang diawasi digunakan untuk mengklasifikasikan setiap gerakan bersama dengan gerakan. Pendekatan baru yang merupakan sistem deteksi botnet yang digunakan untuk mengklasifikasikan kedua aktivitas berdasarkan bot dan yang lainnya adalah komputer klien jaringan.

Ini mencakup bagian halus yang diberi nama M_1 hingga M_5, terkait dengan aktivitas bot di jaringan dan klien terpisah. Sekarang bagian pertama, M_1, adalah modul studi korelasi *Mortal-Growth-Network* (MGN). Modul ini mendeteksi cara jahat dengan melacak metode manusia pada host yang mengacu pada keyboard dan mouse dan menghubungkannya dengan aktivitas jaringan. Sistem menguji perbedaan waktu antara proses yang menghasilkan klik mouse atau kejadian keyboard, pengiriman kejadian tersebut juga memeriksa apakah metode tersebut berjalan di latar depan pada saat itu juga dipertimbangkan.



Gambar 4.6 Pemetaan antara area ML x dan objek y.

Perbedaan waktu yang kecil juga dapat menyiratkan bahwa prosedur tersebut telah dihasilkan dengan bantuan manusia; jika tidak, metode ini dapat ditandai sebagai mencurigakan dan diteruskan ke M_2, M_3, dan M_4. M_2 dan M_3 menggunakan penguasaan yang diawasi untuk mengkategorikan nama domain yang ditanyakan sebagai berbahaya atau jinak dan mengklasifikasikan perilaku jahat pada struktur laptop host masing-masing. M_4 memantau lalu lintas yang dihasilkan melalui cara yang mencurigakan di antarmuka komunitas host.

Paket yang masuk dan harga alternatif antara teknik dan situs web yang jauh dibandingkan. Jika rasio alternatif lebih kecil dari biaya yang telah ditentukan sebelumnya, perilaku bot dicurigai. Akhirnya, setelah setiap modul membuat pilihannya, mesin korelasi - M_6, menggabungkan hasilnya untuk membuat pilihan terakhir menggunakan skema pemungutan suara tertimbang yang ditunjukkan pada Gambar 4.6.

4.6 PEMBELAJARAN TANPA PENGAWASAN

Pembelajaran tanpa pengawasan adalah jenis pembelajaran mesin lainnya, yang termasuk dalam evaluasi struktur yang digunakan untuk menemukan cara menyusun model dalam kumpulan informasi yang sepenuhnya berdasarkan variabel masukan. Tujuan dari pembelajaran tersebut adalah untuk menetapkan fungsi yang merinci pola tersembunyi dalam

fakta yang tidak berlabel. Sasaran utama dari setiap pembelajar ini adalah untuk menyiapkan fitur guna memberikan penjelasan tentang pola tersembunyi dalam fakta yang tidak berlabel. Tidak adanya nilai sasaran (y), atau evaluasi lingkungan eksternal, adalah yang membedakan pembelajaran tanpa pengawasan dari pembelajaran yang diawasi dan penguatan. Jenis pembelajaran tanpa pengawasan yang paling umum disebut pengelompokan. Ini adalah teknik pembelajaran tanpa pengawasan yang digunakan untuk menemukan kesamaan dalam catatan yang tidak berlabel melalui pengelompokannya dalam bagian yang dikenal sebagai kluster.

Melihat bahwa semua faktor informasi tampak sama (tidak berlabel), tujuan dari serangkaian aturan pengelompokan adalah untuk mengenali hubungan di antara setiap faktor fakta dan mengelompokkannya sesuai dengan itu. Dengan cara yang persis sama, karena berkaitan dengan deteksi botnet, proses pengelompokan harus digunakan untuk mengelompokkan lalu lintas situs web daring dengan ciri-ciri serupa dengan tujuan untuk memilih dan menemukan pengunjung situs dengan penyebab jahat. Gadget deteksi botminer mengelompokkan pengunjung situs komunikasi yang serupa dan pengunjung situs web jahat yang serupa dan menjalankan hubungan go-bundle untuk menemukan cloud yang berbagi desain komunikasi yang sama dan gaya minat jahat yang serupa.

Peran Pembelajaran Tanpa Pengawasan

Algoritma kluster K-Means diusulkan untuk deteksi daring. Pendekatan ini menggunakan keterampilan jaringan float dalam jendela waktu yang telah ditentukan sebelumnya. Tujuannya adalah untuk menetapkan lalu lintas situs terutama berdasarkan kesamaan. Kluster dengan kesamaan lebih besar dari ambang batas yang telah ditentukan sebelumnya dapat digolongkan sebagai mencurigakan; akibatnya host yang terkait dengan aliran tersebut dapat ditandai. Teknik lain adalah layanan kerangka nama domain (DNF). Server C&C digunakan oleh bot untuk pencarian DNF. Pemikiran dengan panduan pemanfaatan adalah bahwa bot yang dipisahkan dari botnet yang setara akan menggunakan kontribusi DNF juga. Strategi ini menggunakan kalkulasi pengelompokan x-way untuk ruang asosiasi yang mungkin terkait dengan botnet. Ketiga, perangkat yang menentukan botnet P2P terlepas dari botnet yang saat ini terlibat dalam minat jahat. Penentuan pendekatan ini adalah untuk menemukan bot P2P dengan cara mengidentifikasi gaya komunikasi C&C yang mencirikan bot P2P.

Perangkat pertama-tama mengidentifikasi host P2P kemudian bot P2P mengidentifikasi di antara host tersebut. Metode ini menggunakan keterampilan tingkat drift, mesin mengasumsikan bahwa node P2P membuat banyak aliran keluar yang gagal. Untuk setiap kelompok aliran, ip spot mereka diperiksa dan untuk setiap ip, awalan bgp mereka diperiksa. Jika rentang awalan bgp yang berbeda lebih kecil dari jumlah yang telah ditentukan sebelumnya, mereka diabaikan. Untuk membedakan pengunjung situs web P2P yang sah dari koneksi bot P2P, penulis berasumsi bahwa bot dari botnet yang sama menggunakan protokol dan jaringan P2P yang sebanding. Selain itu, mereka berasumsi bahwa pasangan menjadi bagian dari sumber bot yang memiliki tumpang tindih lebih lama daripada lalu lintas P2P yang valid

Penggunaan aturan pengelompokan x-way, grup meluncur di sepanjang gaya komunikasi yang sama. Ini berisi beberapa aditif di sepanjang tiga tingkatan. Tingkat pertama memiliki unit tampilan video sederhana A dan C yang menampilkan aliran lalu lintas keluar dan dalam secara serial. Tahap lain disiapkan dari pengelompokan permukaan A dan C yang mengelompokkan pengunjung situs, disempurnakan dengan bantuan cara di perangkat tampilan video masing-masing dari tingkat sebelumnya. Hasil dari kluster tersebut kemudian dilampaui ke tingkat 0,33, korelasi sederhana yang membentuk pilihan terakhir tentang host yang mungkin merupakan aspek dari botnet. Dengan cara menggabungkan hasil dari grup sederhana A dan C.

Metode lain adalah "Aliran Kluster" (CF) yang didasarkan pada biaya kesamaan. Pendekatan ini dipecah menjadi tiga bagian, tahap pertama menguji karakteristik, yang kedua, aliran kluster dan 0,33, pilihan botnet. Di dalam level awal, kapabilitas yang berasal dari muatan drift di dalam durasi waktu sebagai vektor 256 D. Pada level ke-2, aliran dikelompokkan menggunakan algoritma pengelompokan pendekatan-k dan pendekatan-x. Kelompok-kelompok tersebut diteruskan ke fase 0,33 di mana kelompok dengan deviasi terendah yang diketahui ditandai sebagai botnet.

Masalah dengan Sistem Deteksi Botnet yang Ada

Banyak upaya telah dilakukan oleh para peneliti untuk mengembangkan kerangka kerja untuk mendeteksi botnet. Masalah telah diidentifikasi dalam struktur deteksi botnet saat ini dan fase di bawah ini menyoroti masalah tersebut dan solusi yang layak untuk mengatasi kekurangannya. Sistem deteksi didasarkan pada perilaku abnormal pengunjung situs jaringan. Gadget tersebut dapat menemukan pengunjung situs terenkripsi dan tidak memerlukan data lapisan aplikasi. Namun, ia tidak dapat mendeteksi percakapan bot IRC pada port yang tidak disukai. Desain sistem deteksi yang diusulkan yang menangani aliran waktu nyata dapat menyelesaikan masalah dengan sistem ini.

Lalu lintas sistem deteksi botnet peer-to-peer yang ada di gateway dipantau dengan menggunakan teknologi penambangan data. Margin sistem terdiri dari itu; ia benar-benar bekerja paling baik dalam lingkungan komunitas area lokal, dan harus dialokasikan ke tingkat ISP (Penyedia Layanan Internet) untuk menemukan botnet P2P dalam jaringan skala besar. Kedua, gaya hidup era NAT membuat gawai sulit untuk menemukan aliran P2P. Para peneliti mengusulkan komunitas berskala besar yang dirancang untuk deteksi botnet yang lebih baik dan lebih kuat.

Spam Preventing System (SPS) untuk deteksi botnet didasarkan pada *support vector machine* (SVM). Sistem menangani pemblokiran email yang tidak diminta dengan cara memisahkan mesin pengguna akhir dengan mesin server yang valid. Kelemahan gawai adalah ia menggunakan kumpulan data yang sangat kecil, dan ternyata tidak diinginkan untuk kumpulan fakta kecil dan diamati tidak diinginkan untuk server email bisnis kecil. Arah penelitian selanjutnya mengusulkan peningkatan panjang dan variasi unit statistik.

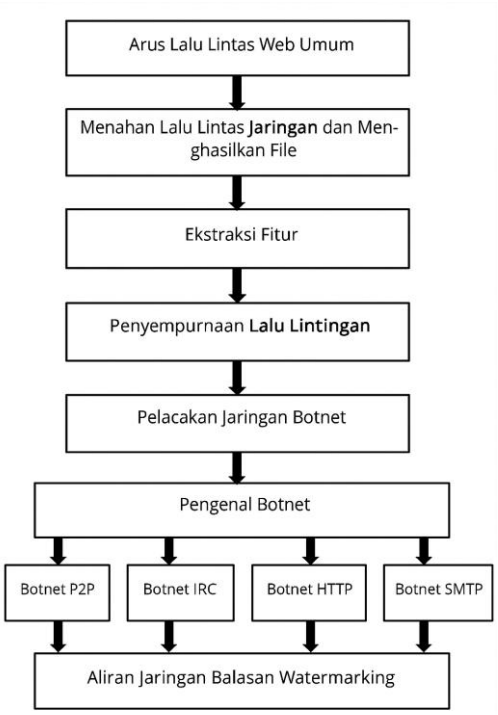
Mengusulkan, mesin deteksi botnet didasarkan sepenuhnya pada mesin yang memperoleh pengetahuan menggunakan data pertanyaan operator panggilan domain. Evaluasi efektivitas pendekatan dengan menggunakan berbagai kumpulan pengetahuan

sistem dan hasil eksperimen menunjukkan bahwa algoritma hutan acak menghasilkan akurasi deteksi dasar yang memuaskan lebih dari 90%. Sistem tidak lagi menangani efek fungsi nama domain pada deteksi, mengusulkan pengujian berkelanjutan dari model yang diusulkan dengan fakta yang lebih besar yang ditetapkan untuk membantu memeriksa efek fitur DNF pada peningkatan akurasi deteksi.Kar1ena masalah sistem deteksi botnet yang ada ini, Sistem Deteksi Botnet Ekstensif (EBDS) telah diperkenalkan.

4.7 SISTEM DETEKSI BOTNET EKSTENSIF (EBDS)

Versi harian yang diusulkan merupakan perluasan dari mesin deteksi botnet yang dibangun. Di dalamnya, lima teknik kelas yang berbeda, khususnya regresi logika yang baik, subruang acak, pengklasifikasi komponen yang dapat diacak, pengklasifikasi multikelas, dan komite acak, telah diterapkan dan dievaluasi dan kemudian pengklasifikasi regresi logistik berubah menjadi yang paling cocok untuk deteksi botnet. Namun, model mereka menjadi salah karena fakta bahwa penggunaan jaringan biasa berkembang secara eksponensial dalam panorama dunia maya dan setelah titik tertentu dalam waktu kerangka kerja mungkin tidak lagi dapat membedakan statistik yang berbahaya dan jinak; akibatnya memberikan dorongan ke atas untuk harga denda palsu yang berlebihan. Namun mereka mengusulkan bahwa memanfaatkan jaringan saraf dengan pengoptimalan penguasaan mendalam akan menghasilkan jawaban yang lebih tinggi dengan cara memperhitungkan botnet yang berkembang dan mengurangi rasio positif palsu.

Oleh karena itu, pemeriksaan ini mengusulkan rekayasa penemuan botnet non-eksklusif yang akan menyampaikan kalkulasi optik, kalkulasi sistem saraf dengan peningkatan pembelajaran mendalam untuk menangani identifikasi botnet dalam lingkungan digital yang berkembang secara eksponensial. Penilaian tingkat pelaksanaan kerangka kerja yang diusulkan dan diperluas (menggunakan Algoritma optik) akan ditetapkan. Kerangka kerja yang diusulkan akan menangani meskipun pengenalan botnet; bukti yang dapat dikenali dari jenis botnet dan manfaat apa yang diberikan botnet. Kerangka kerja pengenalan botnet yang diusulkan juga akan menangkap reaksi mengatur pemeriksaan air lalu lintas. Penandaan air ini akan digunakan untuk melacak kembali ke botmaster untuk menentukan alamat IP botmaster. Korespondensi antara bot dan botmaster bersifat dua arah dan intuitif; Hal ini karena setiap kali botmaster menyampaikan pesan, bot harus menjawab, dan jawaban tersebut harus kembali ke bot master. Model yang diusulkan menargetkan pemberian tanda air pada lalu lintas respons dari bot dengan tujuan agar kita akhirnya dapat melacak kembali ke botmaster yang ditunjukkan pada Gambar 4.7.



Gambar 4.7 Sistem Deteksi Botnet yang Luas (EBDS).

Karena bot ternyata lebih merusak, upaya di zona tersebut meningkat, menciptakan berbagai teknik untuk mengidentifikasi dan melindungi dari botnet. Sejauh ini, pendekatan deteksi berbasis pembelajaran mesin telah terbukti relatif kuat, tetapi masih memiliki keterbatasan. Pengenalan yang tepat waktu, deteksi yang tepat waktu, pelacakan waktu nyata, dan fleksibilitas terhadap ancaman baru adalah masalah yang masih harus dipecahkan. Metode ML khusus memiliki kekuatan dan kelemahan eksklusif seperti yang terlihat dalam posisi yang ditunjukkannya dalam deteksi bot.

Prosedur premis yang terukur (misalnya, penggambaran spekulasi), menekankan dirinya dengan hubungan antara sorotan (x) dan target (y). Sementara dalam perjalanan untuk secara efektif menetapkan arah penggunaan bot, ini harus digambarkan dengan panduan sorotan yang dipilih dan karenanya mengharapkan beberapa informasi eksplisit tentang seperti apa hasil ini. Terutama bergantung pada properti SL, subjek ini telah memanfaatkan keakuratan strategi SL untuk secara tepat menyadari bot berdasarkan beberapa properti yang dikenali dan tepat.

Keakuratan SL mungkin cukup kuat dalam melawan pengunjung bot yang mencoba menyamarkan diri mereka di antara pengunjung situs yang valid, mengingat beberapa ciri khusus dari pengunjung jahat. Dalam survei kami terhadap strategi SL, kami telah menemukan mode yang umum. Selain wawasan khusus tentang lalu lintas bot yang ditemukan di dalam area karakteristik, strategi SL bekerja dengan datar. Strategi SL juga dapat mengalahkan sifat bot yang tertutup. Seperti yang terlihat pada Tabel 4.1, strategi pembelajaran yang diawasi dipekerjakan untuk kasus-kasus di mana beberapa fitur tertentu dipahami.

Bab ini berisi pendekatan deteksi botnet baru yang disebut sebagai Sistem Deteksi Botnet Luas (EBDS). Kelas proses deteksi ini menjanjikan deteksi otomatis; Hal ini mampu menggeneralisasi informasi tentang pengunjung situs komunitas yang jahat dari pengamatan yang tersedia, sehingga menangkal hilangnya prosedur deteksi berbasis tanda tangan yang paling mampu menemukan anomali pengunjung yang diketahui. Untuk memerangi botnet di medan yang dinamis, pekerjaan masa depan pada deteksi botnet, penambahan kemampuan terbaru untuk mengklasifikasikan hasil dari jenis botnet tertentu.

Selain itu, mereka mengusulkan penggabungan baris yang dimodifikasi untuk menentukan IP sumber botmaster. Identitas kehidupan botnet, jenis layanan apa yang dapat diakses botnet juga merupakan usulan untuk pekerjaan masa depan. Otomatisasi seluruh mesin deteksi botnet dapat menghasilkan pengetahuan tentang atribut botnet dan dapat mempermudah pemusnahan botnet dari suatu gawai. Dengan sedikit keberuntungan, gawai yang diusulkan ini dapat memiliki insiden hasil negatif palsu yang lebih rendah; gawai tersebut juga dirancang dengan tujuan untuk mendeteksi botnet secara efektif di suatu komunitas, dan gawai tersebut juga akan memilih jenis botnet itu dan juga mengklasifikasikan layanan dan aplikasi apa yang dapat diakses botnet.

Tabel 4.1 Aspek metode pembelajaran mesin dalam sistem deteksi botnet.

Metode	Wilayah pelacakan	Prospek deteksi	Tingkat positif akurat	Tingkat positif tidak akurat
Terawasi	Khusus	Tanda Tangan (P2P)	98%	2.3%
Terawasi	Khusus	Tanda Tangan (IRC)	Bukan Angka	10% - 20%
Terawasi	Khusus	Tanda Tangan (C&C)	87%	20%
Terawasi	Khusus	Signature	91%	0.56%
Terawasi	Khusus	Signature (IRC)	Not a Number	30%
Tanpa Pengawasan	Klaster	Alam	99%	1%
Tanpa Pengawasan	Klaster	Alam	100%	20%
Tanpa Pengawasan	Klaster	Tanda tangan	95%	4%
Tanpa Pengawasan.	Klaster	Tanda tangan	100%	0.2%
Tanpa Pengawasan.	Klaster	Tanda tangan	95%	Bukan Angka

BAB 5
PEMFILTERAN SPAM MENGGUNAKAN AI

Peningkatan jumlah email yang tidak diinginkan yang disebut spam telah menciptakan kebutuhan akan filter spam untuk mengurangi waktu dan upaya dalam mengelola kotak masuk serta mengelola penyimpanan. Filter spam yang efisien dapat mencegah pengguna dari penipuan dunia maya, dan data pengguna juga dapat diamankan dari spammer. Akhir-akhir ini, metode pembelajaran mesin telah sangat berhasil dalam mendeteksi dan memfilter email spam. Model-model ini pada dasarnya "belajar" dari pengalaman sehubungan dengan beberapa tugas dan mampu menemukan "kesamaan" dalam banyak pengamatan yang berbeda. Studi ini membahas berbagai metode pemfilteran spam menggunakan teknik Kecerdasan Buatan yang ada dan membandingkan kekuatan dan keterbatasannya.

5.1 PENDAHULUAN

Apa itu SPAM?

Email atau surat elektronik merupakan cara yang sangat cepat sekaligus ekonomis untuk bertukar informasi melalui internet. Dalam beberapa dekade terakhir, penggunaan email telah berkembang pesat dan begitu pula penggunaan surat tidak sah, yaitu spam. Spam adalah email yang tidak diminta dan tidak diinginkan dan sering disebut sebagai Surat Massal yang Tidak Diminta (UBM); spam telah menjadi masalah di internet. Masalah di balik email spam adalah bahwa email tersebut umumnya dikirim secara massal, yang menyebabkan pemborosan waktu untuk memfilter sekumpulan email yang diinginkan dari kotak masuk serta mengunci banyak ruang sistem dan menyerap bandwidth komunikasi. Pemfilteran spam adalah teknik untuk mengidentifikasi email spam dan memisahkannya dari email tidak resmi, yaitu email yang berguna, untuk mengurangi waktu dan tenaga. Teknik rekayasa pengetahuan dan pembelajaran mesin telah memperoleh beberapa keberhasilan dalam pemfilteran spam. Di sisi lain, para pengirim spam juga mencari teknik baru untuk melewati filter, termasuk pengaburan kata dan spam gambar.

Tujuan Spamming

Meskipun pengguna terkadang merasa spam menjengkelkan, spam telah menjadi bagian integral dari industri periklanan bagi para pebisnis. Beberapa dekade lalu, tujuan utama pengirim spam adalah menyalahgunakan norma sosial untuk mempromosikan produk mereka, tetapi di era saat ini, peluang tersebut telah diambil oleh penjahat dunia maya, email spam digunakan untuk mengumpulkan informasi penting tentang pengguna atau terkadang membujuk pengguna untuk mengunjungi tautan web yang meragukan. Spam, untuk memenuhi tujuannya, harus memiliki muatan yang dituju dan filter spam harus digunakan untuk melindungi pengguna dari spam.

Input dan Output Filter Spam

Seperti yang telah kita bahas, penyaringan spam adalah teknik untuk memisahkan email spam dari email palsu. Filter menggunakan berbagai cara untuk menyelesaikan tugas

pemisahan. Terkadang Header email digunakan untuk memfilter sementara pemfilteran mungkin berdasarkan konten email, karakteristik pengirim atau target, atau mungkin pengirimnya adalah spammer.

Gambar 5.1 Awan kata dari kata-kata umum dalam email spam.

Gambar 5.1 menunjukkan kata umum yang digunakan di cloud dalam email spam. Seperangkat aturan yang ditetapkan dengan baik diperlukan agar filter dapat melakukan pekerjaan seperti konten email atau informasi statistik yang berasal dari korpus pelatihan atau metode umpan balik untuk menerapkan pengklasifikasi probabilistik. Keluaran paling sederhana yang diinginkan dari filter spam apa pun adalah klasifikasi biner, yaitu, 0 untuk spam yang pasti atau 1 untuk bukan spam. Kategorisasi ordinal yang lebih kompleks dapat mencakup langkah-langkah seperti pasti spam, kemungkinan spam, tidak yakin, kemungkinan bukan spam atau pasti bukan spam.

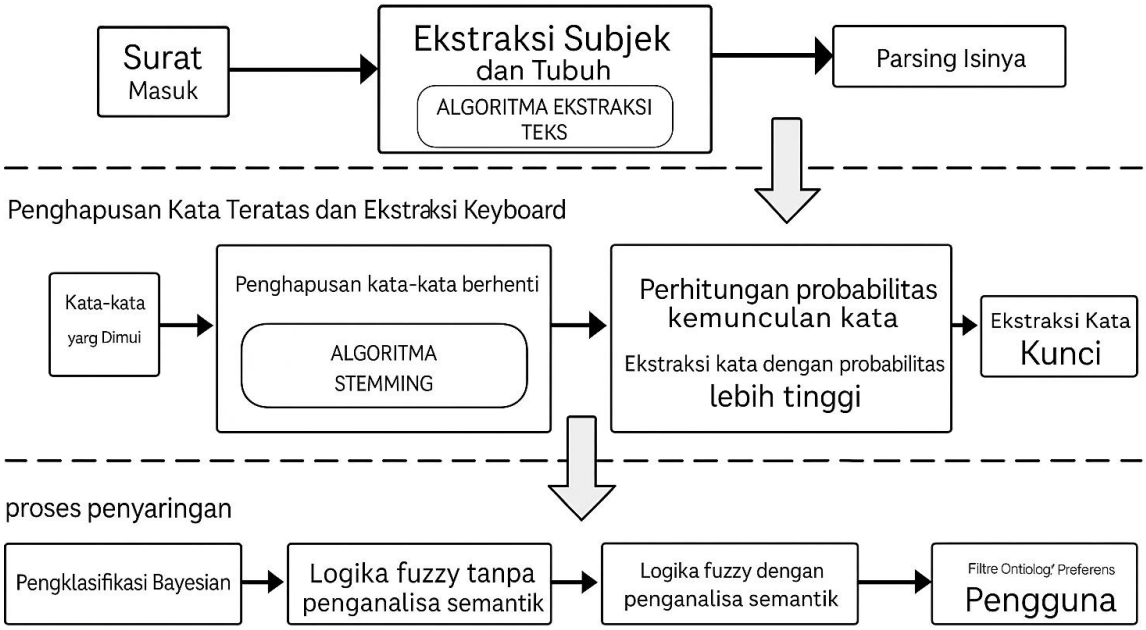
5.2 TEKNIK PENYARINGAN SPAM BERBASIS KONTEN

Penyaring Berbasis Kesamaan Sebelumnya

Penyaruhan semacam itu umumnya juga disebut penyaring berbasis Memori/Instansi. Teknik berbasis pembelajaran mesin ini menggunakan data dari email yang disimpan dan mengklasifikasikan titik instans baru sesuai dengan kesamaan dengan contoh tersimpan sebelumnya atau set pelatihan.

Penyaring Penalaran Berbasis Kasus

Penyaruhan berbasis kasus lebih sering berfungsi sebagai algoritma pembelajaran penguatan. Dalam metode ini, sistem memiliki satu set email/kasus yang sebelumnya diklasifikasikan, bukan satu set instruksi yang tetap. Setiap kali, pada saat klasifikasi, sistem mencoba mencocokkan karakteristik kasus uji dengan set kasus sebelumnya dalam koleksi sistem. Kemudian email uji juga akan disertakan dalam memori sistem sebagai email yang diklasifikasikan untuk klasifikasi di masa mendatang. Beberapa kasus awal dapat diklasifikasikan menggunakan klasifikasi Naïve Bayesian dan kemudian sistem akan menambahkan kasus baru yang ditunjukkan pada Gambar 5.2 (Sharma dan Suryawansi, 2016).



Gambar 5.2 Mekanisme filter spam yang umum.

Filter Email Berbasis Ontologi

Penyaringan spam email berbasis ontologi adalah jenis deteksi spam yang sama sekali berbeda. Teknik ini dikembangkan dengan penggunaan pohon ontologi dan juga dengan evolusi algoritma klasifikasi yang lebih baik. Akhir-akhir ini, keterbatasan utama pendekatan ini adalah antarmuka antara dua sistem independen. Untuk mendapatkan masukan yang diinginkan, prototipe dibentuk yang benar-benar menggunakan informasi ini. Selain itu, keterbatasan utama penyaringan berbasis ontologi adalah diperlukannya pra-proses email masukan dalam format CSV. Selain itu, penyaringan ini juga memiliki keterbatasan dengan klasifikasi Bayesian karena digunakan untuk mengklasifikasikan email yang menggunakan ontologi untuk memahami isinya.

Model Pembelajaran Mesin

Ini adalah ilmu menemukan hal-hal yang tidak diketahui dari data, memperoleh wawasan yang dapat ditindaklanjuti/diprediksi dari data, menciptakan produk data yang berdampak pada bisnis, mengomunikasikan kisah bisnis yang relevan dari data, dan membangun kepercayaan dalam keputusan yang mendorong nilai bisnis. Pembelajaran mesin adalah program yang "belajar" dari pengalaman untuk beberapa tugas. Secara sederhana, model pembelajaran mesin apa pun dapat didefinisikan sebagai representasi realitas yang disederhanakan atau parsial, yang didefinisikan untuk menyelesaikan tugas atau mencapai argumen.

Pembelajaran Terbimbing

Pembelajaran terbimbing hanyalah tugas yang harus dilakukan, untuk mengekstrak deskripsi/pelabelan atau pola dari data, berdasarkan pelatihan. Dalam cabang ini, contoh pelatihan diberi label oleh pengawas (manusia) yang kemudian digunakan untuk memprediksi keluaran untuk contoh selanjutnya. Beberapa aplikasi umum pembelajaran terbimbing meliputi persetujuan kredit, diagnosis medis, dan deteksi penipuan, dll.

Pembelajaran Tanpa Pengawasan

Pembelajaran tanpa pengawasan pada dasarnya adalah menemukan pola/kelompok/kategori yang menarik dalam data berdasarkan bukti. Dalam cabang pembelajaran mesin ini tidak ada data yang diberi label ulang yang digunakan. Kinerja seberapa baik kelompok/pola diukur dari data mentah. Beberapa aplikasi umum pembelajaran tanpa pengawasan adalah segmentasi pelanggan, kategorisasi perilaku pengguna, dan pengelompokan item berdasarkan kesamaan, dll.

Pembelajaran Penguatan

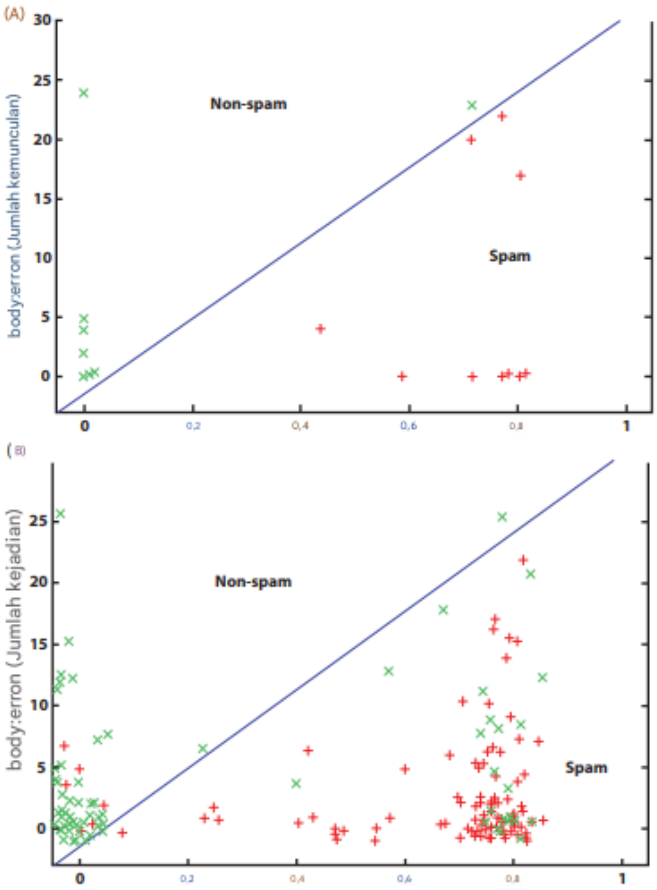
Pembelajaran penguatan adalah bagian terpisah dari Pembelajaran Mesin, yang terkadang juga dikenal sebagai pembelajaran garis besar. Model pembelajaran mesin ini pada dasarnya belajar melalui interaksi coba-coba. Tujuannya adalah untuk memaksimalkan jumlah hadiah yang diterima dari lingkungan. Ini adalah proses interaktif di mana model dilatih melalui interaksi melalui lingkungan dan hadiah diberikan pada setiap keberhasilan. Kinerja diukur berdasarkan jumlah penghargaan yang diterima. Beberapa aplikasi umum lainnya adalah pembelajaran robot dan permainan, dll.

5.3 PENYARINGAN BERBASIS PEMBELAJARAN MESIN

Pengklasifikasi Linier

Pengklasifikasi linier digunakan untuk mengklasifikasikan, katakanlah n fitur yang direpresentasikan dalam bentuk matriks, katakanlah $x[m]$ untuk setiap titik m dalam ruang berdimensi n. Di sini, matriks koefisien didefinisikan sebagai $\beta = (\beta_1, \beta_2 \cdots \beta_n)$ dengan ambang t, maka bidang hiper yang membagi bidang menjadi dua bagian akan mengikuti persamaan $\beta \cdot x = t$. Oleh karena itu, semua titik yang terletak di atas bidang, yaitu, dengan $(\beta \cdot x[m] > t)$ disebut sebagai spam dan yang lainnya sebagai bukan spam.

Algoritme – Untuk secara intuitif menemukan bidang hiper pemisah, jika ada, bobot dapat ditingkatkan atau dikurangi untuk setiap titik di sisi yang salah. Di sini, algoritme mengabaikan contoh-contoh yang diklasifikasikan dengan benar. Sekarang pelatihan dapat dihentikan dengan asumsi bahwa pengklasifikasi yang cukup baik ditemukan di set pelatihan. Metode ini cukup sederhana, efisien, dan bertahap. Di sini, penambahan parameter laju dan margin juga dapat digunakan untuk memengaruhi pelatihan pada kesalahan dan nyaris celaka yang ditunjukkan pada Gambar 5.3. hiperbidang Representasi vektor beberapa pesan dalam set pelatihan contoh yang sedang berjalan. Di sini, pemisah diatur dengan sempurna untuk data pelatihan.

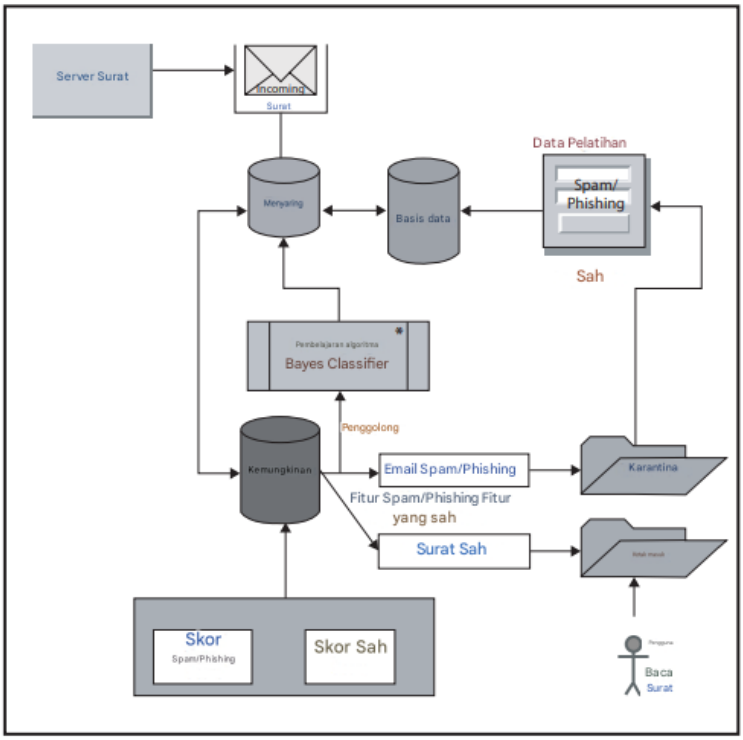


Gambar 5.3 Set hiperbidang untuk data penyamakan.

Di sini hiperbidang yang sama yang diperoleh di atas menghasilkan kegagalan saat digunakan pada seluruh data sampel. Hiperbidang mampu menjaga sebagian besar email spam di sisi spam dan non-spam di sisi lain. Jadi ini dapat dianggap sebagai pengklasifikasi yang wajar tetapi diragukan bahwa ini adalah yang terbaik.

Penyaringan Naïve Bayes

Naïve Bayes juga merupakan teknik pembelajaran terbimbing berdasarkan probabilitas dan statika. Metode penyaringan email ini menggunakan seperangkat aturan adaptif dan seperangkat probabilitas terkait ditetapkan menurut keputusan klasifikasi dan email yang diterima. Setiap email dijelaskan oleh seperangkat atribut dan setiap atribut diberi probabilitas menurut berapa kali email tersebut muncul dalam perangkat pelatihan. Klasifikasi Naïve Bayes untuk menyaring spam menggunakan rumus probabilitas sederhana yang dapat diartikan sebagai (di mana c=spam): *"Probabilitas sebuah pesan menjadi spam, dengan mempertimbangkan fitur-fiturnya, sama dengan probabilitas pesan apa pun menjadi spam dikalikan dengan probabilitas fitur-fitur tersebut muncul bersamaan dalam spam dibagi dengan probabilitas mengamati fitur-fitur tersebut dalam pesan apa pun"* yang ditunjukkan pada Gambar 5.4.



Gambar 5.4 Diagram alir pengklasifikasi Naive Bayes.

Support Vector Machines

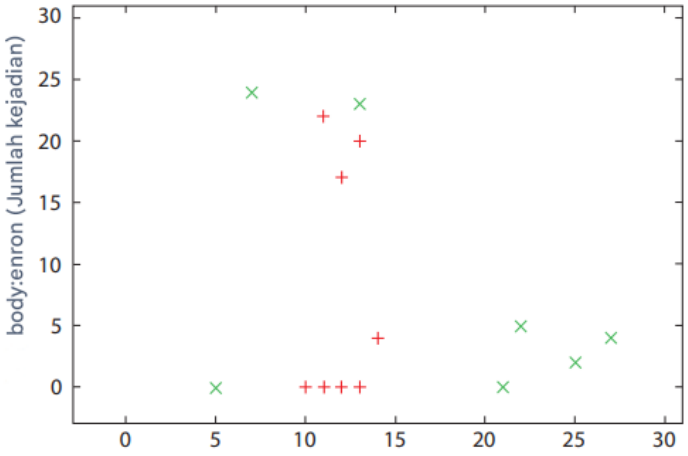
Support Vector Machines adalah algoritma pembelajaran terbimbing yang telah menunjukkan kinerja yang jauh lebih baik daripada pengklasifikasi lain karena batasan multidimensi dan kesederhanaannya. Ia memaksimalkan jarak ke titik contoh terdekat dan titik-titik yang berjarak sama dengan titik contoh yang diberikan disebut sebagai support vectors. Kombinasi linear dari support vectors ini membentuk pengklasifikasi atau hiperbidang pemisah.

Algoritma: Masukkan set pelatihan, katakanlah S, dan tentukan fungsi kernel dalam bentuk $\{c_1, c_2,.....,c_n\}$ dan $\{d_1, d_2,.....,d_n\}$. Tetapkan sejumlah tetangga terdekat, katakanlah k. Kemudian rancang loop for dua tahap, untuk loop luar, tetapkan $c=ci$ dari 1 hingga n. Loop dalam berlaku untuk j dari 1 hingga q, di mana fungsi pengklasifikasi SVM $f(x)$ dirancang dengan parameter penggabungan (c, d). Dengan menggunakan kondisi if-else, fungsi pengklasifikasi $f(x)$ dibandingkan dengan pengklasifikasi terbaik yang diberikan oleh k-fold cross validator. Oleh karena itu, perintah return diberikan untuk mengklasifikasikan pesan sebagai spam atau bukan spam yang ditunjukkan pada Gambar 5.5.

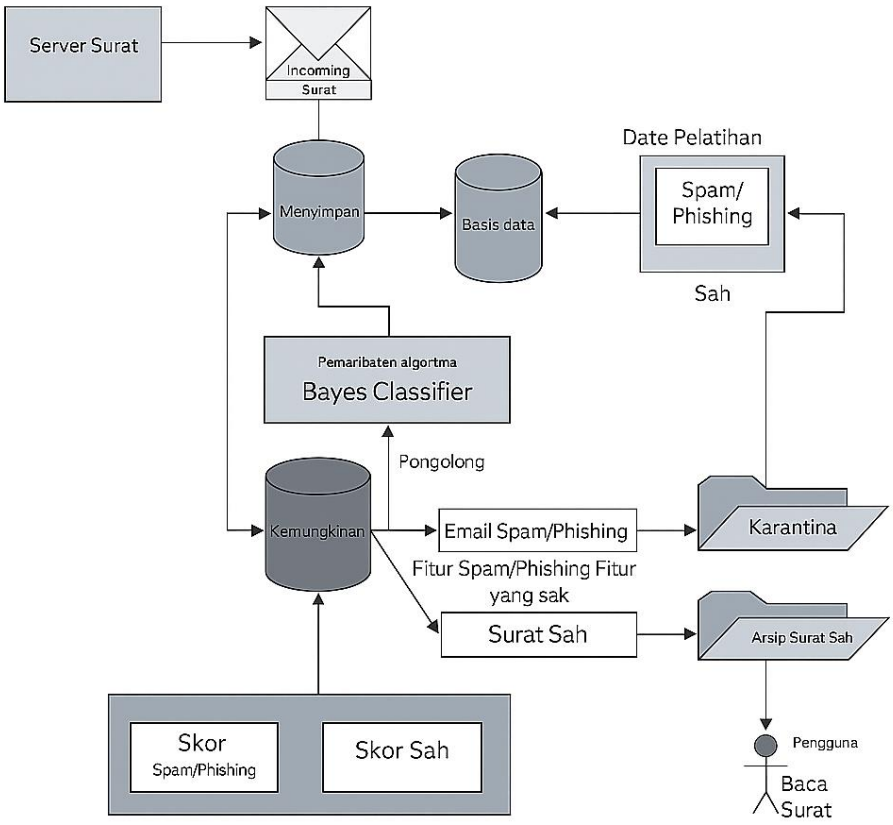
Jaringan Syaraf Tiruan dan Pemfilteran Berbasis Logika Fuzzy

Jaringan syaraf tiruan adalah sekelompok unit pemrosesan sederhana yang didasarkan pada jaringan syaraf biologis. Setiap unit terhubung dengan tetangganya dengan beberapa bobot yang ditetapkan. Masing-masing menerima masukan dari satu tetangga dan mengirimkan keluaran yang dimodifikasi ke tetangga lainnya. Secara umum, jaringan syaraf tiruan dirancang dengan menghubungkan tiga lapisan, yaitu lapisan masukan, lapisan tengah, dan lapisan keluaran.

Logika fuzzy juga merupakan pendekatan metodis untuk mengklasifikasikan email sebagai spam atau sah secara otomatis dengan mempertimbangkan konten email.



Gambar 5.5 Hasil simulasi pengklasifikasi SVM.



Gambar 5.6 Arsitektur penyaringan spam email berbasis logika fuzzy.

SIstem ini dapat secara otomatis beradaptasi dari konten email dan membangun basis datanya sendiri. Pengklasifikasi akan dibangun dari set pelatihan email yang telah diklasifikasikan

sebelumnya. Arsitektur sistem penyaringan spam email dibangun dalam tiga langkah, yaitu, pra-pemrosesan, pembuatan set pelatihan, dan klasifikasi yang ditunjukkan pada Gambar 5.6.

Pra-pemrosesan:

Pra-pemrosesan data merupakan langkah terpenting sebelum menerapkan logika apa pun ke set pelatihan. Email juga perlu diproses terlebih dahulu sebelum menggunakannya untuk pelatihan dan klasifikasi guna membuat algoritme lebih efisien serta untuk memperoleh hasil yang optimal. Langkah ini meliputi pembersihan teks dengan menghapus semua tag HTML beserta menyingkirkan kata-kata yang tidak penting seperti tanda baca, dsb. Kemudian stemming dilakukan untuk mengurangi kelangkaan matriks akhir kata-kata dalam bentuk T = <t1, t2... tN> di mana N adalah jumlah total kata penting yang sekarang dapat disebut sebagai token. Sekarang jumlah kemunculan setiap token dalam setiap kategori, c termasuk spam, sah}, ditentukan.

Pembuatan Set Pelatihan:

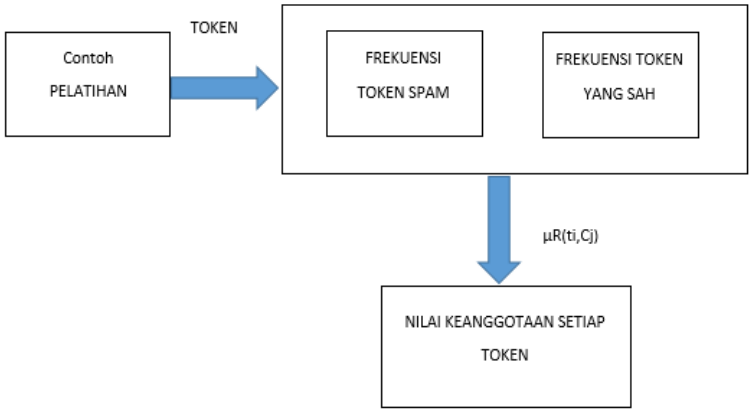
Langkah ini dilakukan untuk membangun model berdasarkan karakteristik set pelatihan email yang telah ditentukan sebelumnya. Set pelatihan dipilih dengan mempertimbangkan berbagai konten dan subjek. Sekarang setelah menjalani fase pra-pemrosesan, kita dapat memperoleh (fi,c) yang menunjukkan frekuensi token t_i apa pun, dalam kategori 'c'. Dari kumpulan data ini, jaringan saraf dapat dibangun dengan bobot terhitung dari setiap token untuk membentuk matriks T x C dan relasi kategori token fuzzy tersebut digunakan untuk menetapkan nilai relasi, yaitu, R: T × C → [0, 1]. Di mana, μR(ti,cj) - Derajat token ti apa pun, dalam kategori Cj. fi cj - Frekuensi kemunculan token t_i dalam kategori tertentu Cj fi, sah +fi, spam - Frekuensi kemunculan token ti di semua Kategori

Klasifikasi:

Oleh karena itu, algoritma ini didasarkan pada relasi fuzzy antara setiap email yang diterima dan frekuensi setiap token yang dikandungnya. Jadi untuk menentukan kategori setiap email yang diterima, katakanlah d, frekuensi setiap token dalam d dapat ditentukan menggunakan

$$\mu_d(t_i) = \frac{f_{i,d}}{max_{t_j \in d}\{f_{i,d}\}}$$

Kemudian operator konjungsi fuzzy dan disjungsi fuzzy dapat digunakan untuk mengukur rasio kesamaan dan nilai ambang batas untuk menghitung kesamaan fuzzy yang ditunjukkan pada Gambar 5.7.



Gambar 5.7 Diagram blok untuk pembuatan set pelatihan.

Analisis Kinerja

Korpus contoh email dapat diambil dan dimasukkan ke dalam Penyaringan Berbasis Header, Penyaringan Berbasis Konten, atau teknik penyaringan URL apa pun. Di sini, kami juga mempertimbangkan satu set contoh email dan setelah melalui berbagai pengklasifikasi/metodologi penyaringan, perbandingan kuantitatif berikut dicatat dalam hal tingkat Akurasi model, tingkat penarikan kembali, dan nilai presisi.

a. Filter Email berbasis Ontologi - Dalam sekitar 5-6 kali percobaan, akurasi model ini bervariasi antara 0,95 hingga 1 yang menunjukkan bahwa peningkatan jumlah masukan menyebabkan penurunan akurasi. Nilai penarikan kembali berfluktuasi dari 0,94 hingga 1 sementara nilai presisi berada di antara 0,97 hingga 1.
b. Pengklasifikasi Linier – Set data yang berbeda digunakan untuk klasifikasi teknik ini. Akurasinya menjadi 97,72% menggunakan algoritme yang dijelaskan di atas. Tingkat recall-nya tercatat sebesar 97,29% dan presisi sebesar 97,55%.
c. Naïve Bayes Filtering – Model ini memiliki akurasi yang cukup rendah, yaitu antara 0,78 hingga 0,8 karena peningkatan jumlah input memerlukan lebih banyak pelatihan. Recall berada pada kisaran 0,85 hingga 0,91 dan presisi berfluktuasi dari 0,77 hingga 0,91 yang menunjukkan basis pengetahuan yang digunakan untuk melatih model tidak mencukupi.
d. Support Vector Machines – Classifier ini memiliki tingkat akurasi sebesar 0,96 yang menunjukkan classifier cukup baik untuk digunakan untuk penyaringan. Tingkat recall-nya mencapai 95% dan nilai presisi sebesar 93,12%.
e. Jaringan saraf dan pemfilteran berbasis logika fuzzy – Sekali lagi, tingkat akurasi untuk model ini berkisar antara 0,8 hingga 0,9 pada berbagai lintasan pada kumpulan data yang menunjukkan penurunan nilai akurasi karena peningkatan jumlah masukan karena pelatihan yang tidak memadai. Tingkat penarikan kembali bervariasi antara 0,87 hingga 0,91 sementara presisi berkisar antara 0,84 hingga 0,97.

Dalam pembahasan di atas, kami telah meninjau beberapa algoritme pembelajaran mesin yang paling populer dan koheren yang digunakan untuk memecahkan masalah email spam. Studi ini membuat analisis kinerja yang jelas dan perbandingan berbagai teknik penyaringan spam berbasis konten. Eksperimen menunjukkan perbandingan kuantitatif dari teknik yang tercantum dalam hal nilai Akurasi, Ingatan, dan Presisi. Sebagian besar model pembelajaran mesin dibangun di atas data pelatihan dan karenanya akurasinya bervariasi. Meningkatkan jumlah masukan membutuhkan lebih banyak pelatihan, yaitu, basis pengetahuan yang besar dan memadai yang berisi sebagian besar variasi data.

Set pengujian yang lebih besar menunjukkan akurasi yang lebih rendah karena set data yang tidak mencukupi selama fase pelatihan. Pengklasifikasi linier telah menunjukkan akurasi yang baik tetapi lebih spesifik untuk set data masukan sementara mesin vektor pendukung secara umum telah menunjukkan kinerja yang memuaskan. Pendekatan Bayesian dalam deteksi spam email telah memberikan dasar yang baik untuk membuat pengklasifikasi Meta-spam. Diperlukan penelitian lebih lanjut mengenai jaringan saraf dan sistem berbasis logika fuzzy untuk menemukan cara yang jauh lebih efisien untuk penyaringan spam, tidak hanya dalam hal penyaringan berbasis konten tetapi juga deteksi gambar spam.

BAB 6
PERAN AI DALAM KEAMANAN SIBER

Pengetahuan kecerdasan buatan (AI) dihasilkan dan disajikan oleh mesin. Setiap program yang mampu mempelajari pola dan mengambil langkah untuk meningkatkan kemampuannya dalam menyelesaikan suatu tugas dapat dianggap sebagai bentuk AI. Studi tentang AI berkembang sangat pesat, mencakup berbagai bidang seperti pembelajaran mendalam (*deep learning*), pengenalan bahasa alami, pemrosesan ucapan, pemecahan masalah dinamis, interaksi manusia-mesin, hingga penalaran semantik dan pemrosesan informasi naratif. Kemajuan AI memungkinkan siapa pun mengakses kemampuan setara dengan para ahli, karena sistem yang dibangun dapat terus belajar melalui penggunaan berulang. Hasilnya, sistem tersebut bisa memberikan jawaban yang semakin akurat, bahkan berpotensi melampaui keahlian manusia dalam bidang tertentu.

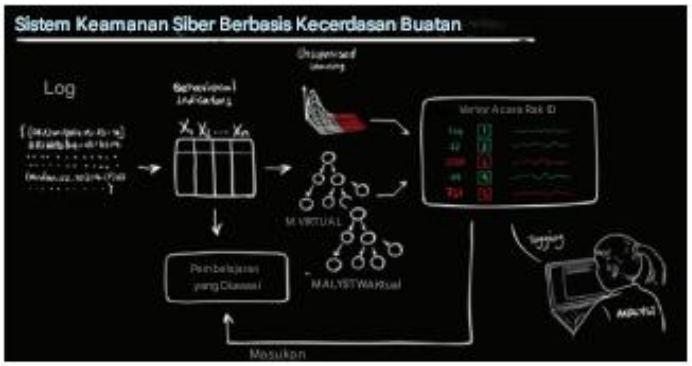
Seiring meningkatnya kecerdasan mesin, AI akan mampu memahami informasi kompleks yang berkaitan dengan manusia. Melalui pemanfaatan data sensor digital, AI dapat digunakan untuk mengembangkan konsultan cerdas, instruktur virtual, maupun mitra kerja digital. Namun, kemajuan ini juga membawa risiko dan tantangan, seperti ancaman terhadap privasi, keamanan data, moralitas penggunaan, dan pengelolaan sistem. Studi ini berfokus pada risiko-risiko tersebut, khususnya dalam konteks keamanan siber, serta bagaimana AI dapat digunakan untuk mendeteksi, menganalisis, dan menangani ancaman dalam jaringan digital.

6.1 PENDAHULUAN

Awalnya, Kecerdasan Buatan (AI) merupakan sebuah ide untuk meniru otak manusia, dan untuk mengeksplorasi masalah dunia nyata secara holistik. Kecerdasan berbasis komputer paling populer untuk aplikasi film dan akademisnya; AI memungkinkan untuk menyediakan banyak informasi dan menangani informasi tersebut dengan bijak. Kecerdasan buatan telah digunakan untuk memberikan aplikasi cerdas dalam berbagai zona, misalnya, pertahanan atau eksplorasi ruang angkasa. Situs-situs ini memiliki sejarah yang kaya akan berbagai solusi pemecahan masalah. Kemudian, AI melihat aplikasi di bidang layanan medis. AI telah digunakan untuk hal-hal seperti diagnostik, saran pengobatan, dan perawatan hati-hati.

Teknologi kinerja dapat dicirikan sebagai kecerdasan buatan yang menyediakan alat untuk memecahkan masalah yang kompleks dan menegangkan. Kecerdasan buatan adalah kombinasi dari inovasi data dan kecerdasan fisik, yang dapat digunakan secara elektronik untuk mencapai tujuan. Kebijaksanaan adalah kapasitas untuk berefleksi dengan membangun ingatan dan menerima, melihat contoh, menyusun pilihan yang kuat, dan memperoleh fakta. AI dapat membuat mesin bekerja seperti manusia, tetapi mereka dapat bekerja lebih cepat dan lebih lembut. Sebagian besar kelompok pada level keamanan ini mengendalikan area keamanan. Standar pengaturan yang digunakan dalam penelitian ini dapat dianggap sebagai

klasifikasi ilmiah. Kami memperkenalkan beberapa area keamanan siber yang ditunjukkan pada Gambar 6.1:



Gambar 6.1 Sistem keamanan siber.

- ✓ Keamanan struktur
- ✓ Keamanan akhir
- ✓ Keamanan aplikasi
- ✓ Keamanan IoT
- ✓ Keamanan web
- ✓ Keamanan dan respons kejadian
- ✓ Ancaman intelijen
- ✓ Keamanan seluler
- ✓ Keamanan cloud
- ✓ Manajemen kepemilikan dan akses
- ✓ Keamanan sistem
- ✓ Keamanan tenaga kerja

Investigasi ini meneliti presentasi yang memanfaatkan penalaran buatan manusia dari berbagai pembuat. Aplikasi yang dipertimbangkan memanfaatkan kesadaran buatan manusia untuk meramalkan dan membedakan bahaya dan kegagalan keamanan data. Percakapan aplikasi dimaksudkan untuk memberikan diagram pengaturan keamanan jaringan yang memanfaatkan kecerdasan buatan yang dapat diakses, dan apa yang dapat mereka tawarkan untuk menentukan solusi untuk masalah tersebut.

Keamanan siber:

Langkah-langkah perlindungan digital terkait dengan manajemen bahaya, penangkapan bahaya, dan kekuatan kerangka kerja. Tema eksplorasi utama menggabungkan prosedur yang diidentifikasi dengan pengenalan perilaku organisasi yang berbahaya dan malware, serta pertanyaan TI yang diidentifikasi dengan keamanan TI. Jadi, perlindungan jaringan dapat dikarakterisasikan sebagai berbagai aktivitas yang diambil untuk mencegah serangan digital dan akibatnya serta menggabungkan penggunaan langkah-langkah penangkalan penting. Perlindungan jaringan didasarkan pada pemeriksaan ancaman terhadap suatu tautan atau instalasi. Struktur dan bagian dari prosedur keamanan jaringan asosiasi dan rencana pelaksanaannya bergantung pada bahaya yang dapat diukur dan investigasi bahaya.

Tujuannya adalah untuk menetapkan berbagai prsedur dan aturan keamanan otoritatif. Ancaman yang secara langsung atau tidak langsung ditujukan pada proyek publik atau swasta muncul dari dalam atau luar batas negara. Keadaan ancaman adalah sudut ancaman yang berisi data tentang ancaman kecelakaan dan serangan kapal. Melalui penyalahgunaan atau kekurangan, ancaman menyebabkan kerugian atau realokasi properti.

Gambar 6.2 Memberdayakan analis keamanan.

Faktor penting adalah perencanaan dasar untuk memerangi ancaman, dan perlindungan yang memadai terhadap dampak berbahaya dari ancaman. Tindakan untuk memerangi ancaman digital dapat ditingkatkan dengan meningkatkan dasar-dasar keamanan jaringan, memperluas pengetahuan setiap orang tentang bahaya, meningkatkan pelaksanaan dan pemeliharaan keamanan. Kuncinya adalah mengenali tantangan keamanan jaringan dan bereaksi dengan tepat. Bagian penting dari perlindungan jaringan adalah kemampuan untuk menahan kapasitas agar dapat bekerja di bawah serangan digital, memiliki pilihan untuk segera mengakhiri serangan dan membangun kembali tugas-tugas jaringan ke keadaan normal sebelum serangan. Undang-undang yang sesuai dan diskusi yang mendalam dan tepat diperlukan untuk mengatasi masalah ini. Cara-cara untuk memerangi serangan siber dapat dibahas secara rinci seperti yang ditunjukkan pada Gambar 6.2.

6.2 PENGATURAN KORESPONDENSI PERLINDUNGAN DAN KEAMANAN DIGITAL

Kecerdasan buatan tidak hanya menimbulkan ancaman dan bahaya; kecerdasan buatan juga dapat berfungsi sebagai solusi untuk masalah. Penyaluran informasi dan arahan operasional digunakan untuk mengenali, mencegah, dan mengidentifikasi serangan siber. Jadwal keamanan data harian dibuat oleh manusia atau dimekanisasi. Aktivitas diagnostik analitis ini bergantung pada instruksi yang dibuat oleh pakar keamanan TI, yang mengabaikan serangan yang tidak sesuai dengan aturan yang ditetapkan. Strategi yang disusun secara

mekanis dapat menghasilkan hasil positif palsu, menciptakan rasa tidak percaya yang umum terhadap sistem, dan membutuhkan upaya manusia untuk memeriksa kasus.

Keselamatan Operasi dan Respons Peristiwa

Kesadaran buatan manusia adalah salah satu pilihan yang tepat yang secara masuk akal dapat mencegah hilangnya banyak nyawa seperti yang dibuat oleh program gelar Online Bosses in Security di Eastern Kentucky College. Kekuatan otak buatan manusia untuk gambaran Reaksi Bencana (AIDR) selama aktivitas krisis akan digunakan oleh pusat aktivitas krisis.

AI2

Prosiding Pertemuan Global Kedua Belas tentang Keamanan dan Keselamatan Militer ICCWS2019 IMIT *Computer Discipline and the Artificial Intelligence Laboratory* (CSAIL) dan Patrice telah membuat tahap mata-mata AI2 untuk meramalkan serangan digital. Menurut Conner-Simons, tahap AI2 memiliki opsi untuk mencapai ketepatan 86% dalam mengidentifikasi serangan digital, yang beberapa kali lebih unggul dari pemeriksaan sebelumnya. Untuk mencegah serangan, AI2 mengenali tindakan yang meragukan menggunakan kalkulasi bagian informasi menggunakan kalkulasi AI. Mencegah serangan, Peneliti juga menambahkan model tahap (bacaan terarah) dalam koleksi informasi yang menyertainya, yang memungkinkan pembelajaran lebih lanjut. Program ini juga siap untuk terus membuat representasi baru dalam hitungan jam, yang pada dasarnya dapat meningkatkan kecepatan kemampuan serangan digitalnya.

CylanceProtect

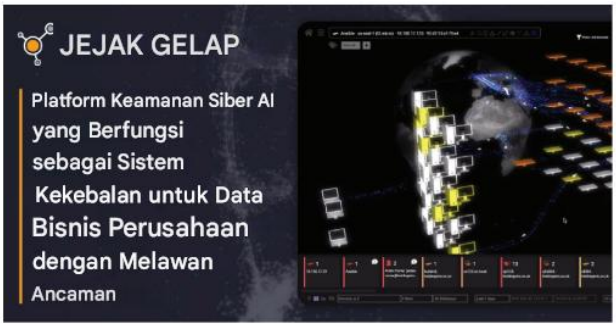
CylanceProtect adalah alat keamanan data terpadu, yang menggabungkan keunggulan kecerdasan buatan dan panel keamanan data untuk menghindari kontaminasi malware. Panel keamanan data digunakan untuk memastikan terhadap serangan skrip, serangan memori, atau gadget eksternal. Tidak seperti peralatan keamanan tradisional yang bergantung pada pemeriksaan investigasi dan perilaku klien dalam mengidentifikasi bahaya alami, CylanceProtect:

- Menggunakan kecerdasan buatan (bukan tanda tangan) untuk mengenali dan mengamankan perangkat lunak berbahaya yang dikenal dan tidak dikenal yang berjalan pada perangkat akhir
- Menghindari serangan zero-day yang dikenal dan tidak dikenal
- Menjaga gadget tanpa mengganggu klien akhir

6.3 PELACAKAN HITAM

Pelacakan hitam adalah organisasi keamanan data, yang dapat membantu membedakan dan mengidentifikasi bahaya digital yang berkembang yang dapat mencegah jaminan data konvensional. Pelacakan hitam menggunakan inovasi Sistem Kekebalan Perusahaan (EIS) dan menggunakan kalkulasi AI dan standar numerik untuk mengidentifikasi kelalaian di dalam organisasi data asosiasi. EIS menggunakan teknik statistik, yang berarti tidak perlu mengeksploitasi tanda atau petunjuk, dan dapat menyoroti serangan perlindungan digital yang tidak umum. EIS dapat mengenali dan bereaksi terhadap beberapa bahaya digital yang direncanakan secara menyeluruh, termasuk bahaya internal yang tercakup dalam

organisasi data. Dengan menggunakan instrumen pembelajaran dan pengukuran, EIS dapat menyesuaikan dan kemudian mengetahui bagaimana setiap klien, gadget, dan jaringan data bertindak, sehingga dapat mengidentifikasi perilaku yang mencerminkan bahaya digital yang sebenarnya.



Gambar 6.3 Keamanan siber AI.

Inovasi eksplorasi redup memberi organisasi perspektif menyeluruh tentang organisasi data dan memungkinkan mereka untuk bereaksi lebih tegas terhadap bahaya dan mengurangi bahaya yang ditunjukkan pada Gambar 6.3. Ia melihat sistem info dan memungkinkannya untuk merespons ancaman dengan lebih kuat dan mengurangi risiko. Alih-alih menggambarkan perilaku "buruk" di masa lalu dan mengandalkan metode serangan sebelumnya, pembacaan mesin jejak gelap, dan perspektif peluang Bayesian, dapat secara otomatis memodelkan dan mengintegrasikan data dengan kekuatan dan kecepatan. Jalur hitam memantau data mentah, seperti integrasi layanan cloud, yang dikirimkan melalui jaringan secara real time, tanpa gangguan, seperti, operasi bisnis dan koneksi. Ia juga memberikan opini langsung dari semua peristiwa digital dengan menulis serangan berkelanjutan atau berbahaya.

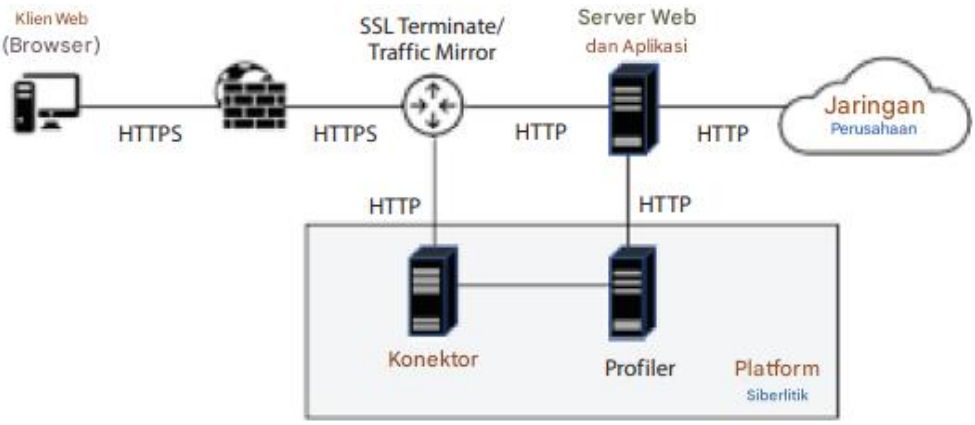
Jalur hitam melibatkan empat mesin matematika, yang menggunakan banyak metode matematika, seperti duplikasi persamaan Bayesian. Tiga model pertama menghasilkan model komunikasi untuk individu dan perangkat yang mereka gunakan, serta kelompok secara keseluruhan. Ketika kinerja abnormal terdeteksi, satu atau lebih mesin ini akan menampilkan pesan ke pemisah bahaya, yang fungsinya adalah untuk mendeteksi keadaan serius dan laporan malfungsi yang tidak dapat diperiksa dengan benar. Pengelompokan pendekatan Basest yang konsisten menghasilkan akurasi gangguan yang tepat pada skala organisasi.

Keamanan Web

Cyberlytic Profiler

Cyberlytic Profiler menggunakan berbagai prosedur baru untuk menganalisis hasil yang tidak menentu dari lalu lintas web untuk mengurangi aksesibilitas tanda standar dan inovasi otoritas. Teknik-teknik ini membantu mengidentifikasi bahaya digital yang terus berkembang dan kompleks dengan mengurangi kebutuhan akan mediasi manusia atau upaya terkoordinasi. Profiler menggunakan program AI elektronik dan acak untuk menyelidiki aliran informasi. Perhitungan belajar mandiri gratis menghasilkan perilaku web standar dengan memutuskan pilihan sendiri. Judul yang dapat diakses berfluktuasi dari pekerja web dan menggabungkan

gaya dan teknik sesekali. Sorotan bidang menggabungkan panjang dan apropriasi karakter yang ditunjukkan pada Gambar 6.4.



Gambar 6.4 Cyberlytic Profiler

Dengan mencetak aplikasi web, Profiler dapat memutuskan apakah aplikasi yang diperkenalkan berasal dari distribusi aplikasi yang biasa di wilayah tertentu dari aplikasi web. Teknik ini menentukan apa yang "normal" dalam kelompok tertentu. Sebagai keputusan dapat ditarik untuk memutuskan motif, membuat lalu lintas tidak disukai. Dengan menyorot kelemahan, gangguan yang paling mungkin terjadi dapat ditekankan; yang melanggar hukum dapat dievaluasi untuk bahayanya. Profiler menggunakan teknik terlindungi alternatif untuk membedakan sorotan serangan dari jenis serangan ini: SQL infusion, compose site (XSS) dan Bash. (Cyberlytic.)

Amazon Macie

Amazon Macie adalah fasilitas keamanan informasi yang memanfaatkan AI. Kesadaran kecerdasan buatan memberi Macie alat identifikasi untuk mengidentifikasi, mengatur, dan melindungi data sensitif dari *Amazon Web Services* (AWS). Macie melihat catatan sensitif sebagai informasi khusus atau hak cipta. Demikian pula, ia dapat mengamati bagaimana materi berhak cipta, misalnya, dokumen, disalin, didistribusikan, atau dilihat. Macie memiliki tampilan dasbor yang memutuskan bagaimana informasi digunakan atau dihapus. Aplikasi terus mengamati penggunaan dan ketidakkonsistenan informasi dan memberikan peringatan atau peringatan terperinci jika data menjadi sasaran penggunaan yang tidak sah atau kebocoran dokumen yang tidak diinginkan yang ditunjukkan pada Gambar 6.5.

Amazon Macie adalah layanan keamanan data yang menggunakan AI. Kesadaran kecerdasan buatan memberi Macie perangkat bukti pembeda untuk mengenali, mengatur, dan melindungi informasi sensitif dari *Amazon Web Services* (AWS). Macie melihat data sensitif sebagai informasi khusus atau hak cipta. Demikian pula, ia dapat mengamati bagaimana materi berhak cipta, misalnya, formulir, disalin, didistribusikan, atau dilihat. Macie memiliki dasbor yang memilih bagaimana data digunakan atau dihapus. Aplikasi terus memantau penggunaan dan perubahan data dan memberikan alarm atau peringatan terpisah jika informasi

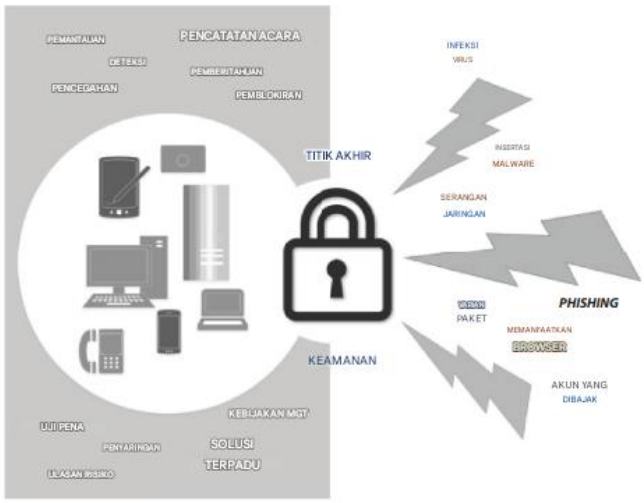
bergantung pada penggunaan yang tidak disetujui atau kebocoran informasi yang tidak diinginkan.

Macie juga dapat berulang kali mengidentifikasi bahaya informasi bisnis, jika informasi didistribusikan ke luar grup tanpa izin, atau jika data tersebut diakses secara tidak sengaja. Macie juga dapat mengidentifikasi bahaya data bisnis, jika data diambil alih di luar asosiasi tanpa persetujuan, atau jika informasi tersebut diakses secara tidak sengaja yang ditunjukkan pada Gambar 6.5.

Gambar 6.5 Amazon Macie.

Perangkat lunak Deep Sensibility dirancang untuk menjaga ponsel dan area layanan dari serangan jahat yang dikenal dan tidak dikenal secara aktual; produk ini bergantung pada penyalahgunaan organisasi saraf palsu. Dengan teknologi instalasi, Deep Instinct dapat mengenali perangkat lunak pada ponsel dan stasiun kerja layanan. Dengan memanfaatkan keterampilan membaca mendalam yang tepat, perangkat lunak ini mampu mengantisipasi serangan siber anonim yang ditunjukkan pada Gambar 6.6 dan Gambar 6.7.

Gambar 6.6 Sensitivitas mendalam.



Gambar 6.7 Keamanan ENDPOINT.

Perancang Deep Sensibility telah menggunakan kalkulasi pembelajaran tingkat lanjut dalam penerapan sistem mereka, yang memungkinkan mereka untuk mendeteksi struktur yang digunakan dalam perangkat lunak jahat. Deep Instinct dapat mengidentifikasi dan mencegah peningkatan pemrograman yang berbahaya di semua tingkat asosiasi. Dengan memanfaatkan kemampuan pemahaman internal dan eksternal, arsitek Deep Instinct membangun organisasi saraf menyeluruh dalam kondisi lab dan melatihnya dengan serangkaian besar pengujian kode jahat. Data tersebut digunakan dengan dokumen berbahaya dan tidak berbahaya untuk menunjukkan organisasi saraf ini. Hasilnya adalah model yang akurat, yang dapat dikirim ke perangkat untuk memastikan dukungan yang berkelanjutan.

Pengetahuan tersebut terdiri dari mendapatkan perangkat lunak untuk mengenali kombinasi aplikasi dan aplikasi yang berbasis pada perangkat lunak berbahaya. Metode pembelajaran Deep Instinct memotong contoh kode perangkat lunak dengan teks yang sangat kecil, untuk menguji perangkat lunak berbahaya. Prosiding Pertemuan Global Kedua Belas tentang Keselamatan dan Keamanan Militer ICCWS2019. Teknik ini seperti urutan genomik, yang juga berisi puluhan ribu seri yang lebih kecil. Fragmen sampel ini dimasukkan ke dalam jaringan saraf untuk menunjukkan jaringan untuk menargetkan penyelesaian. Jenis sistem ini menjalankan kalkulasi yang besar dan sulit, dan rangkaian GPU digunakan untuk membantu kalkulasi ini. Daya kalkulasi GPU jauh lebih cepat daripada CPU. Hasilnya adalah jaringan saraf yang cepat dan matematis yang melibatkan sedikit kontrol komputer, dan dapat digunakan untuk mendeteksi perangkat lunak berbahaya.

6.4 SPARK COGNITION DEEP MILITARY

Spark Cognition Deep Armor mampu mengidentifikasi dan melindungi dari risiko malware, worm, virus Trojan, dan daftar virus, memanfaatkan strategi numerik, misalnya, AI dan koreksi bahasa normal. Konfigurasi Deep Armor menggabungkan sedikit operator dari awal hingga akhir yang dikoordinasikan dengan mesin berbasis cloud, serta kebijakan deteksi

ancaman. Operator akhir mengenali dan menghalangi proyek berbahaya dan bahaya tingkat signifikan lainnya, meskipun ada tanda. Spesialis dimaksudkan untuk memastikan pelanggan, pekerja, gadget portabel dan IoT; Spark, mesin pemahaman berbasis cloud untuk penjelasan Deep Armor menggunakan filter level baru yang ditunjukkan pada Gambar 6.8.

Gambar 6.8 Mesin pemahaman berbasis cloud.

Proses Mendeteksi Ancaman

Lapisan keamanan utama melakukan investigasi dokumen, sama seperti papan manajemen risiko dan penggunaan, untuk mengenali file yang diketahui atau baru dengan cepat. Setelah memfilter catatan file yang dipulihkan, Deep Armor menggunakan kalkulasi intelektual untuk memeriksa dokumen yang tidak jelas dan bentuk ancaman untuk setiap file. Pada tahap berikutnya, mengenali ancaman dan dukungan berbasis cloud memberikan alat penanganan bahasa alami (NLP). Deep NLP tidak hanya memahami bukti daring, tetapi juga memahami konteks seputar ancaman; Deep Armor dapat mengenali material berbahaya seperti yang ditunjukkan oleh kasus abnormal.

Gambar 6.9 Proses mendeteksi ancaman.

Inovasi Deep NLP SparkCognition berfungsi agar banyak halaman data relevan yang diidentifikasi dengan berbagai ancaman dianggap sebagai alat untuk menggabungkan teknologi Deep NLP Spin Cognition. Informasi ini digunakan untuk membedakan ancaman itu sendiri. Alat NLP juga menguji Internet untuk mencari bukti bahaya, yang darinya ringkasan bukti akan dibuat. Sasaran kalkulasi risiko juga akan ditentukan berdasarkan faktor risiko yang diketahui. Akhirnya, sinopsis analisis risiko dapat dibuat dari data yang dihasilkan yang dapat digunakan untuk merencanakan prosedur strategi dan mengatasi masalah yang paling berlaku yang ditunjukkan pada Gambar 6.9.

Vectra Cognito Networks

Vectra Cognito Networks menggunakan kesadaran buatan manusia untuk memberikan gambaran terperinci tentang serangan digital yang sedang berlangsung. Cognito menggabungkan kemajuan pembelajaran mekanis yang disempurnakan, misalnya, pembelajaran top-down dan organisasi saraf, dan model pembelajaran konstan, untuk mengenali penyerang yang tersembunyi dan tidak dikenal dengan cepat dan berhasil sebelum mereka menyebabkan kerusakan.

Cognito juga menghilangkan apa yang disebut “area buta” dengan membedah semua kerangka kerja keamanan dan verifikasi informasi dan aplikasi SaaS untuk lalu lintas jaringan dan dokumen log. Ini memberikan diagram lengkap tentang status klien dan gadget IoT yang diidentifikasi dengan langkah-langkah yang bekerja di cloud dan server farm, mencegah penyerang menutupi apa yang ditunjukkan pada Gambar 6.10.

Lalu lintas jaringan. Log sistem, autentikasi, dan SaaS. IoC (STIX).	Pembelajaran mesin. Analisis perilaku. Efek jaringan.	Triase dan korelasikan ancaman terhadap host. Prioritaskan host berdasarkan risiko. Temukan kampanye Serangan.	Antarmuka pengguna intuitif dengan konteks yang kaya. Aktifkan respons. otomatis Integrasi firewall, titik akhir, SIEM dan NAC.

Gambar 6.10 Vectra AI.

Prosiding Pertemuan Global Kedua Belas tentang Keamanan dan Keselamatan Militer ICCWS2019, Vectra Cognito menggunakan inovasi observasi dan pembelajaran elektronik yang tidak terkontrol, misalnya pembelajaran top-down dan organisasi saraf, untuk melawan serangan digital dan menargetkannya. Kerangka kerja keamanan data asli berupaya

mengidentifikasi serangan digital melalui pencarian tanda-tanda yang diketahui dan disalahgunakan secara pasti. Seorang penyerang digital dapat menggunakan data ini untuk melawan kerangka kerja. Cognito mempelajari tindakan jaringan selama beberapa jangka waktu yang tidak ditentukan, misalnya, hari, minggu, atau bulan.

Cognito mengenali perilaku penyerang digital dalam setiap rangkaian serangan digital. Perilaku penyerang yang diamati diurutkan dan dibandingkan dengan perilaku klien yang umum, menggunakan pekerja yang baru saja diadili karena bahaya. Orang-orang yang penting untuk setiap kampanye serangan digital tertulis diidentifikasi sebagai cerminan perilaku para penyerang. Dalam situasi ini, para pemimpin dapat fokus untuk mengarahkan aset mereka ke serangan paling berisiko terhadap bisnis.

Risiko ancaman:

IBM QRadar Counsellor with Watson menggunakan talenta cerdas IBM Watson (yaitu, penalaran terkomputerisasi) tahap QRadar Safety, tahap yang ditujukan untuk pemeriksaan keamanan data, mengungkap bahaya tersembunyi dan membuat pemeriksaan bahaya. Kerangka kerja secara otomatis mengidentifikasi petunjuk bahaya, menggunakan pengetahuan untuk memanfaatkan kapasitas psikologisnya untuk mendapatkan data sensitif dan akhirnya mempercepat pola reaksi terhadap bahaya keamanan. Qonad Advisor with Watson juga menggunakan fitur Keamanan Siber Watson untuk meneliti dan bereaksi terhadap bahaya keamanan data.

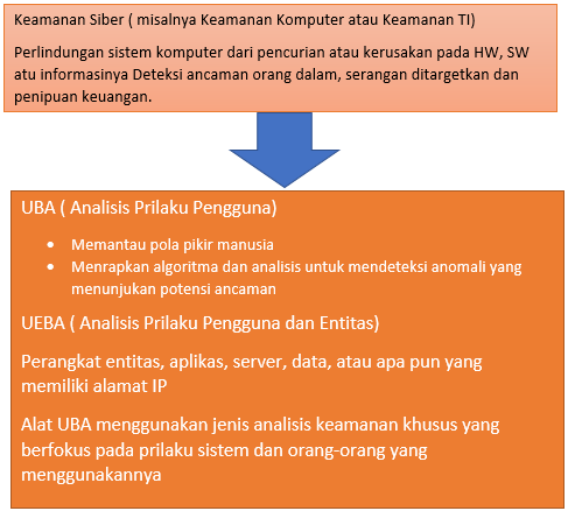
Gambar 6.11 QRadar Advisor.

QRadar Counsellor dengan Watson bekerja melalui langkah-langkah berikut yang ditunjukkan pada Gambar 6.11:

- Ketika tahap Q Radar Safety Intelligence mengidentifikasi bahaya data keamanan, seorang ahli keamanan data dapat merujuknya ke Q Radar Advisor dan Watson untuk pemeriksaan tambahan. Panduan memulai penyisiran total bahaya data keamanan dengan menambang informasi dari perangkat lunak QRadar lokal. Produk tersebut

kemudian akan menggunakan Watson's Cyber Security untuk memimpin penyelidikan intensif terhadap bahaya tersebut

- Watson's Cybersafety mengumpulkan informasi dari berbagai sumber, misalnya, situs web, pertemuan keamanan data, dan rilis berita, dengan cara yang berarti bagi orang tersebut. Terakhir, produk tersebut memerlukan data tambahan tentang data keamanan yang diidentifikasi dengan dokumen berbahaya dan alamat IP yang meragukan.
- Terakhir, penasihat yang berpusat pada QRadar dan Watson.



Gambar 6.12 Keamanan Siber/UBA/UEBA.

User Behavior Analytics (UBA) telah mendapatkan banyak perhatian dalam domain keamanan data. Ketika membangun keamanan perusahaan terhadap bahaya dari luar, asosiasi melindungi diri mereka dari bahaya yang diharapkan di dalam asosiasi. Bahaya dapat dibuat, misalnya, oleh seorang perwakilan atau pemain eksternal, yang dapat menyebabkan kerugian di wilayah tertentu karena perilaku lalai. Bahaya tersebut merupakan ujian karakter dan dapat benar-benar merugikan aset organisasi asosiasi, melemahkan aset tidak berwujud dan kepercayaan konsumen serta merusak reputasi atau nama baik organisasi yang ditunjukkan pada Gambar 6.12 dan Gambar 6.13.

Gambar 6.13 AI dalam Keamanan Siber.

Bab ini mengidentifikasi 12 area penting dalam keamanan siber. Selama penelitian, data dikumpulkan pada 11 solusi kecerdasan buatan. Hasil ini dibagi menjadi beberapa area: Keamanan struktur, keamanan menyeluruh, keamanan web, keamanan dan akuntabilitas, ancaman mata-mata, keamanan seluler, dan keamanan tenaga kerja. Solusi untuk masalah ini terletak pada penyebaran ancaman keamanan siber yang meluas.

BAB 7
PRIVASI AI MULTI-TENANCY

Teknologi multi-penyewa (*multi-tenancy*) semakin populer dan berkembang pesat, terutama dalam sistem komputasi awan dan layanan berbasis platform. Dalam sistem ini, beberapa penyewa (*tenant*) berbagi sumber daya perangkat lunak dan perangkat keras secara bersamaan. Meskipun efisien, sistem ini memunculkan tantangan serius dalam hal privasi dan keamanan, karena semua data penyewa disimpan dalam satu basis data bersama, namun harus tetap terisolasi dan hanya dapat diakses oleh masing-masing penyewa.

Masalah utama dalam multi-penyewa adalah bagaimana menjaga agar data tetap aman, tidak bocor antar penyewa, dan tetap terlindungi dari potensi serangan siber. Untuk menjawab tantangan tersebut, diterapkan teknologi kecerdasan buatan (AI) guna memperkuat sistem privasi dan keamanan. AI bekerja layaknya pikiran manusia atau hewan yang cerdas ia menganalisis pola, belajar dari data, dan melakukan penyesuaian untuk mencapai tujuan sistem secara optimal. Dalam konteks ini, AI digunakan untuk mendeteksi anomali, menganalisis risiko secara real-time, dan mengelola kebijakan akses serta perlindungan data secara adaptif dan otomatis. Dengan demikian, kerangka kerja multi-penyewa yang diperkuat oleh kecerdasan buatan dapat menciptakan lingkungan yang lebih aman, efisien, dan andal, tanpa mengorbankan fleksibilitas dan skalabilitas sistem.

7.1 PENDAHULUAN

Sangat penting untuk membahas tantangan baru yang muncul dalam lingkungan multi-tenancy. Penggunaan internet yang luas saat ini memastikan bahwa banyak penyewa pengguna berbagi data dan perangkat lunak yang sama untuk aplikasi umum. Basis data juga umum untuk setiap penyewa dalam sistem umum, tetapi masalah utamanya adalah mengelola privasi dan keamanan data semua departemen; dengan penyewa departemen yang berbeda sangat sulit untuk dikelola. Jadi dengan menggunakan konsep kecerdasan buatan, kita harus meningkatkan masalah privasi dan keamanan. Dengan menggunakan kecerdasan buatan, pertama-tama kita harus memahami masalah kompleksitas privasi dan keamanan. Kemudian kita dapat menganalisis berbagai utas dan kompleksitas dalam keamanan dan dengan cepat mengidentifikasi solusi dari masalah tersebut.

Dengan menggunakan kecerdasan buatan, kita menemukan hubungan antara penyewa dan kebutuhan umum akan sumber daya. Dengan menggunakan konsep kecerdasan buatan, konsep privasi dan keamanan dapat ditingkatkan. Seiring dengan meningkatnya penggunaan teknologi, demikian pula risiko privasi dan keamanan. Dengan menggunakan kecerdasan buatan, isolasi setiap area akses data penyewa harus dipertahankan. Multi-tenancy sangat populer di pasar teknologi saat ini. Ada sejumlah besar penyewa yang menggunakan perangkat lunak umum atau platform tunggal sehingga tidak perlu mengembangkan banyak perangkat lunak untuk banyak penyewa sehingga biayanya juga berkurang. Konsep multi-tenancy sangat fleksibel untuk dipelihara, dikembangkan, dan dibagikan oleh banyak penyewa sekaligus. Jadi

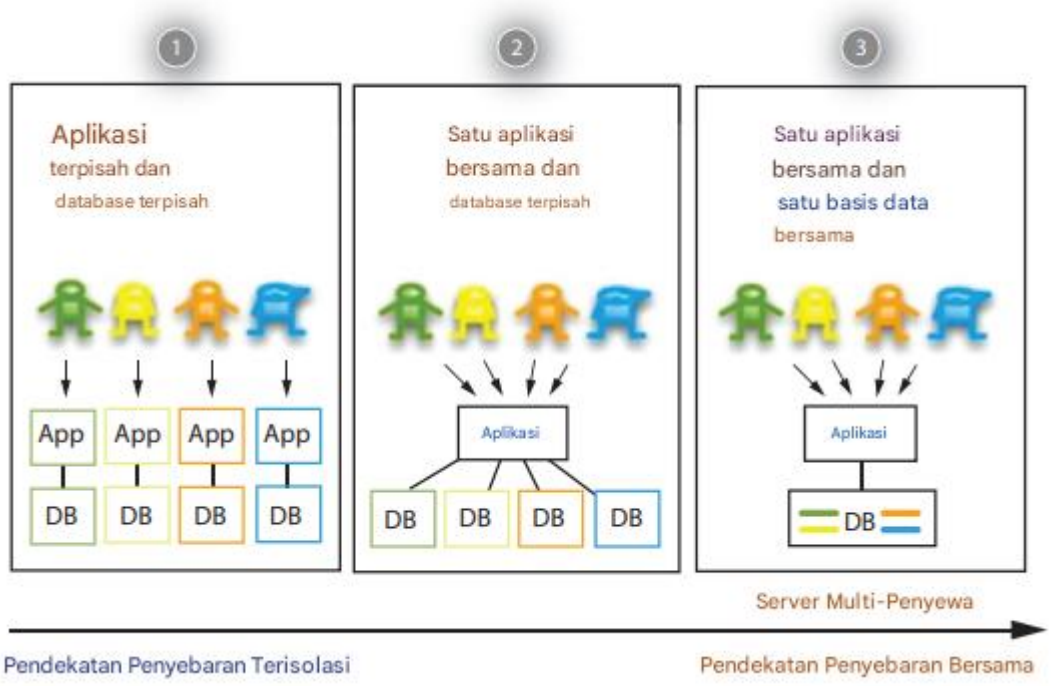
konsep privasi dan keamanan harus dikembangkan dengan cepat menggunakan kecerdasan buatan.

7.2 KERANGKA MULTI-TENANCY

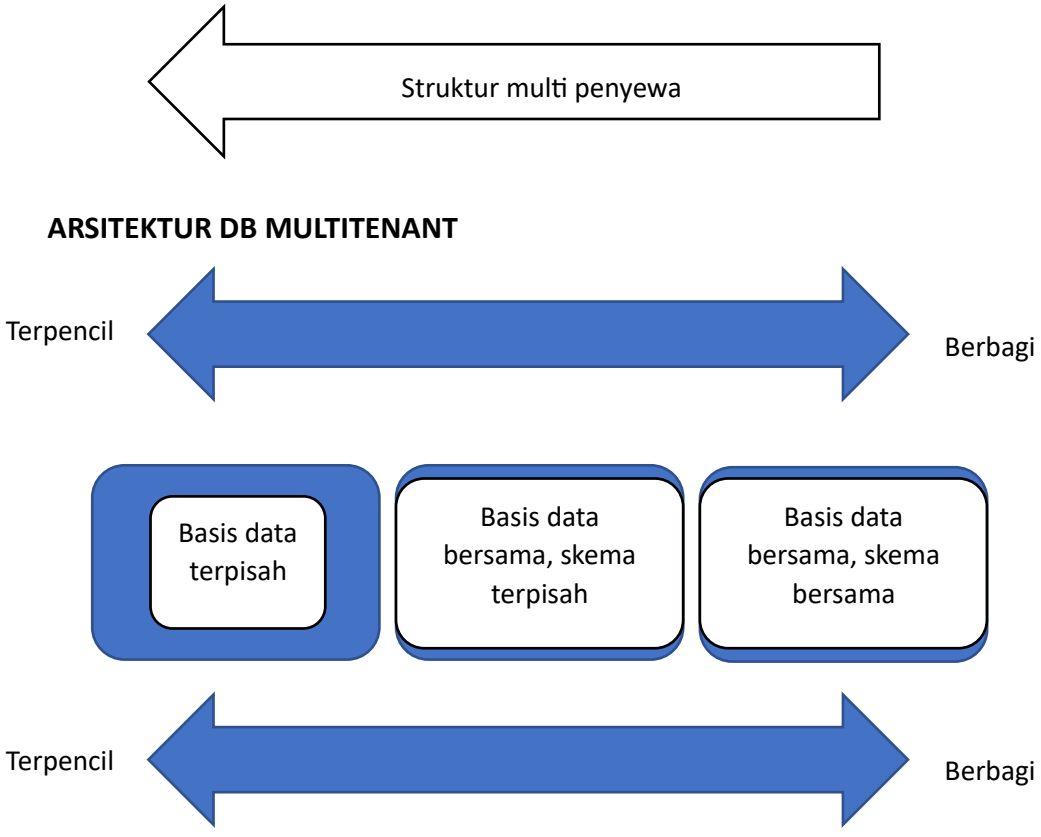
Arsitektur sistem multi-tenancy adalah blok basis data yang dapat dibagikan dengan pengguna atau pelanggan atau klien yang dikenal sebagai penyewa; penyewa menggunakan blok basis data untuk proses atau untuk pekerjaan yang diperlukan yang dapat dilakukan oleh penyewa. Banyak penyewa dapat melakukan pekerjaan mereka menggunakan modul atau blok kode sebagai pekerjaan atau tugas yang diperlukan. Setelah menyelesaikan tugas, mereka dapat mengubah tugas atau dapat mengubah kebutuhan basis data. Konsep kunci utamanya adalah bahwa basis data tidak diubah; hanya bagian dari basis data yang dapat digunakan oleh banyak penyewa sesuai dengan kebutuhan mereka. Mereka dapat memperbarui aplikasi tempat mereka bekerja tetapi tidak mengubah kode atau basis data yang merupakan elemen dasar organisasi.

Semua konten konsep atau submodul menonjolkan konsep kecerdasan buatan. Saat ini, lebih banyak perusahaan dan organisasi yang menggunakan konsep struktur sistem multi-penyewa. Untuk pertumbuhan organisasi dan tujuan komputasi, yang merupakan kebutuhan zaman sekarang untuk melakukan pekerjaan yang cepat dan akurat, pengembangan lebih lanjut dengan pengguna atau penyewa yang tepat dapat melakukan pekerjaan dengan organisasi yang sama di lokasi yang berbeda. Basis data saat ini memiliki banyak data, banyak elemen seperti Tabel, daftar, hubungan antara tabel, kueri, dan banyak lagi elemen dalam tabel. Jadi untuk menjaga aliran data yang tepat dan keamanan serta isolasi data dengan banyak pengguna atau penyewa, konsep kecerdasan buatan diperlukan.

Isolasi data merupakan tantangan utama dalam sistem multi-penyewa; data dapat dibagikan oleh lebih banyak pengguna untuk menjaga pekerjaan dengan cara yang tepat tetapi data harus dipisahkan dalam banyak penyewa. Pemisahan data dapat dilakukan dalam tiga kondisi dasar yang diberikan dalam Gambar 7.1 dan 7.2. Untuk membuat implementasi dengan konsep isolasi data. Mengisolasi basis data sangat penting untuk digunakan oleh banyak pengguna basis data yang sama menggunakan konsep basis data Terpisah. Semua penyewa dapat berbagi banyak sumber daya, basis data, atau blok kode untuk memenuhi persyaratan organisasi.



Gambar 7.1 Struktur multi-tenancy.

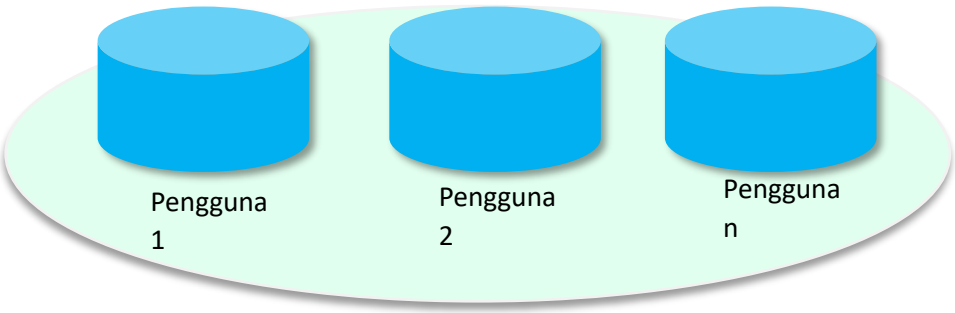


7.3 KEAMANAN DATA MULTI-TENANT BERBASIS AI

Dalam konsep multi-tenant, keamanan dan privasi merupakan tantangan besar untuk mengelola semua data dengan konsep menggunakan kecerdasan buatan menjadi lebih mudah untuk dipelihara, dan biaya juga harus dikurangi menggunakan algoritma AI. Semua pengguna

atau organisasi menuntut keamanan yang lebih karena mereka semua menggunakan basis data yang sama. Pekerjaannya berbeda tetapi data sama-sama aman, dan privasi merupakan persyaratan utama pengguna. Untuk tujuan itu mereka membayar sejumlah uang dan juga menuntut tingkat keamanan yang tinggi sehingga semuanya dapat dilakukan dengan menggunakan semua konsep AI yang baru. Dalam kecerdasan buatan, pengembang dapat melakukan banyak pekerjaan terkait isolasi data yang digunakan algoritma untuk menjaga isolasi data di basis pengguna atau penyewa.

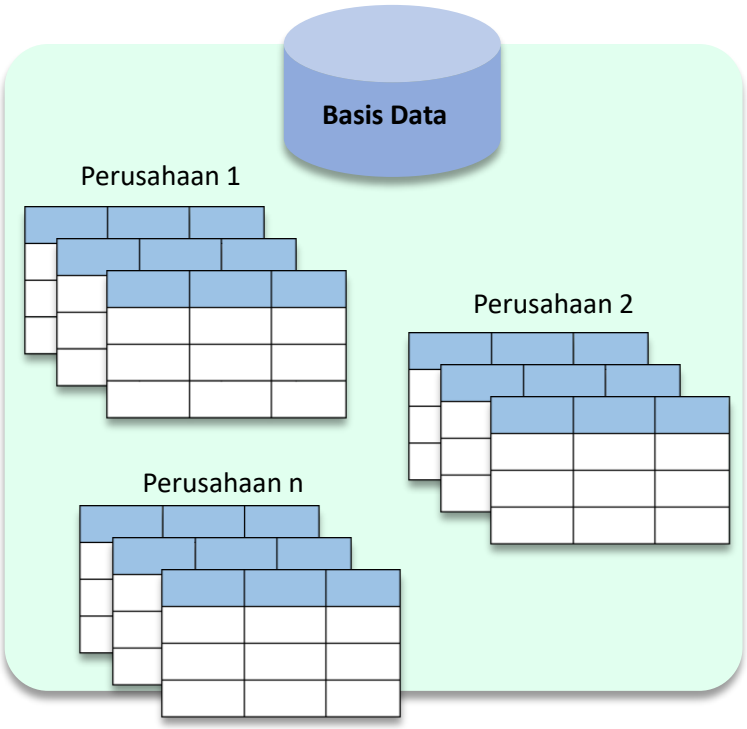
Sering kali penyewa menggunakan tabel yang sama atau basis data yang sama untuk pekerjaan yang terpisah dan juga menyimpan pekerjaan di tabel umum. Jadi sangat penting untuk mengisolasi data dan menjaga privasi dan keamanan di antara semua penyewa. Tujuannya adalah untuk bekerja secara efisien tanpa kebocoran data atau masalah peretasan dan menghasilkan pekerjaan yang cepat dan efisien.



Gambar 7.3 Multi-pengguna dalam sistem.

Seperti yang ditunjukkan pada Gambar 7.3, basis data umum dapat digunakan bersama oleh banyak pengguna atau perusahaan atau beberapa penyewa. Mereka berbagi tabel dan data umum, skema umum, dan struktur umum. Jadi, keamanan tinggi diperlukan untuk mengelola data dan memastikan bahwa data tersebut tidak dapat diakses oleh penyewa yang salah. Dengan menggunakan kecerdasan buatan, skema yang berbeda dibuat untuk setiap penyewa guna berbagi basis data umum tetapi menggunakan skema yang berbeda untuk membuat pekerjaan lebih aman dan terjamin.

Dalam kecerdasan buatan, konsep ini memungkinkan banyak atau beberapa penyewa untuk membuat area terpisah yang diketahui dalam basis data; mereka menggunakan data umum tetapi bekerja secara terpisah dan juga menyimpan pekerjaan mereka dalam basis data yang sama setelah selesai. Penyewa lain juga melakukan hal yang sama jika penyewa yang berbeda yang bekerja dalam modul basis data umum juga dapat melakukan hal yang sama; dengan menggunakan konsep AI (kecerdasan buatan), pekerjaan beberapa penyewa akan dipisahkan atau diamankan. Jadi, dengan menggunakan konsep kecerdasan buatan, isolasi data dengan cara yang aman dan juga skema yang terpisah dapat menciptakan kemampuan bagi setiap penyewa untuk bekerja secara efisien menggunakan algoritme AI yang ditunjukkan pada Gambar 7.4.



Gambar 7.4 Beberapa perusahaan dalam satu basis data.

Melanjutkan tiga konsep keamanan basis data dalam sistem multi-penyewa menggunakan kecerdasan buatan, yang pertama adalah memisahkan data dan juga memisahkan penyewa. Yang kedua, data umum dibagikan tetapi struktur dan skema dipisahkan. Namun, pada struktur dan skema ketiga keduanya umum dan menggunakan konsep kecerdasan buatan, kebijakan keamanan dan privasi dipertahankan. Jadi menurut skema, algoritme AI bergantung dan menggunakan algoritme tersebut, biayanya bergantung, menurut penggunaan algoritme dalam sistem multi-penyewa. Ada risiko tinggi peretasan data oleh penyewa lain sehingga konsep kecerdasan buatan dapat membuat semua struktur dan skema lebih aman dan pengawasan privasi juga lebih kuat dan lebih mampu memuaskan penyewa.

Seperti yang kita semua tahu, data sangat penting bagi semua organisasi. Menurut keamanan dan akses yang aman, sistem multi-penyewa diadopsi oleh organisasi. Semua kontrol tidak dapat dikelola oleh organisasi dan juga tidak dapat dikelola oleh penyewa itu sendiri, sehingga untuk itu, kecerdasan buatan diperlukan untuk menjaga keamanan dan semua skema data yang dapat dikelola disiapkan oleh konsep kecerdasan buatan untuk membuat basis data lebih aman dan lebih mudah dikelola. Juga menjaga biaya kebijakan keamanan dan privasi, karena kita tahu keamanan adalah tantangan besar dalam sistem multi-penyewa. Kecerdasan buatan dapat memenuhi permintaan besar konsep privasi dan keamanan dalam sistem multi-penyewa.

Jadi lingkungan yang aman dituntut dalam sistem multi-penyewa yang baik. Dan sistem multi-penyewa yang aman adalah kebutuhan saat ini. Lingkungan yang aman diperlukan untuk berbagi struktur basis data. Dalam basis data, hubungan antara tabel dan tabel bersama, skema bersama, data bersama harus diisolasi antara beberapa penyewa. Dan pekerjaan yang

dilakukan oleh penyewa harus aman dan terjamin, mengirimkan dan memperbarui basis data oleh banyak penyewa paralel, semua pekerjaan harus dilakukan dalam lingkungan yang aman dan terjamin dalam sistem basis multi-penyewa yang disediakan oleh algoritma basis konsep kecerdasan buatan.

Kecerdasan buatan mengembangkan aplikasi dunia nyata untuk pemeriksaan keamanan. AI menyediakan skema komputasi yang kuat dan terjangkau untuk membuat struktur multi-penyewa lebih aman dan privasi penyewa harus dipertahankan. AI menyediakan layanan yang baik untuk tingkat pemeliharaan penyewa dari layanan keamanan yang baik dan memprioritaskan dukungan dalam skema yang dikembangkan oleh AI. Di masa depan, jutaan skema pengembangan untuk kebijakan keamanan dan privasi akan dikembangkan dalam konsep kecerdasan buatan menggunakan algoritma kecerdasan buatan.

Pekerjaan Terkait

Dalam bab ini, pekerjaan yang terkait dengan konsep kebijakan privasi dalam kerangka kerja multi-tenancy menggunakan konsep kecerdasan buatan telah diteliti. Konsep kecerdasan buatan digunakan untuk mengembangkan sistem multi-tenancy dan membuatnya aman dan terlindungi, terisolasi, dengan privasi dan daya tahan untuk masa depan. Dengan penggunaan kecerdasan buatan, peretasan basis data dan penipuan transisi dapat ditemukan dengan sangat cepat. Biaya pemeliharaan dan kebijakan keamanan dan privasi akan dikurangi dengan menggunakan konsep kecerdasan buatan. Telah banyak pekerjaan yang dilakukan dalam banyak bab atas dasar privasi dan keamanan sistem multi-tenancy menggunakan konsep kecerdasan buatan.

Data bab ini berguna untuk menemukan parameter fungsional dan non-fungsional sistem multi-tenancy sehubungan dengan algoritma keamanan kecerdasan buatan. Dalam bab ini, kami telah membahas secara rinci konsep privasi, keamanan, isolasi data, daya tahan, dan pemeliharaan struktur sistem multi-tenancy. Kami telah memberikan perincian tentang konsep privasi dan keamanan kerangka kerja sistem multi-tenancy, dan berbagai konsep penggunaan sistem multi-tenancy sesuai dengan persyaratan dan pemeliharaan keamanan dan konsep privasi menggunakan kecerdasan buatan.

Dengan penggunaan konsep kecerdasan buatan, kita dapat menjaga faktor keamanan, kinerja, biaya, fleksibilitas sistem multi-tenancy. Bab ini merupakan kontribusi terhadap keamanan dan privasi dalam penggunaan sistem kecerdasan buatan pada sistem multi-tenancy. Bab ini digunakan untuk menemukan solusi maksimal guna membuat kebijakan keamanan dan privasi dalam kecerdasan buatan dalam sistem multi-tenancy lebih aman. Dengan memahami sistem menggunakan kecerdasan buatan, literatur ini digunakan untuk memahami dan menemukan persyaratan sumber daya dan layanan serta metode privasi untuk mengembangkan berbagai layanan seperti waktu respons, beban jaringan, dan layanan manajemen throughput.

Dalam sistem multi-tenancy, konsep keamanan dan privasi menggunakan algoritma kecerdasan buatan sangat berguna dan membuat basis data lebih aman serta memungkinkan pekerjaan yang aman. Penggunaan algoritma AI atau konsep AI memberikan kemampuan penggunaan ulang dan struktur yang aman untuk setiap sistem multi-penyewa. Jadi data harus

akurat, aman, benar, dan tahan lama. Setiap sistem multi-tenancy ingin bekerja dalam privasi dan keamanan untuk bekerja dan mempertahankan daya tahan. Jadi, menggunakan konsep AI menyediakan lingkungan yang aman untuk melakukan pekerjaan dan mempertahankan aliran kerja dalam sistem. Ini menyediakan lingkungan yang baik dan anggun.

Jika basis data tidak aman, penyewa tidak lagi dapat memelihara basis data dan strukturnya. Dalam lingkungan saat ini di mana banyak orang bekerja dari jarak jauh di banyak lokasi berbeda, konsep kecerdasan buatan sangat penting. Untuk penerapan ulang sistem multi-tenant dan keamanan yang diperlukan dalam sistem multi-tenant, sangat penting untuk menggunakan konsep kecerdasan buatan yang benar untuk memberikan keamanan dan privasi ke basis data utama dalam kerangka kerja multi-tenant.

Jadi, konsep kecerdasan buatan digunakan untuk mengembangkan sistem multi-tenant guna memastikan bahwa kinerja, keamanan, privasi tidak hilang dan daya tahan dipertahankan.Konsep multi-tenancy bukanlah proses yang mudah untuk mempertahankan isolasi data, privasi, keamanan, dan banyak konsep lainnya. Untuk menggunakan konsep kecerdasan buatan, konsep keamanan dan privasi harus dikembangkan. Dengan bantuan kecerdasan buatan, sistem multi-penyewa menyediakan kebijakan privasi yang aman dan berkembang, yang membuat bekerja di berbagai basis data menjadi lebih stabil. Penyewa yang berbeda bekerja di basis data yang sama dalam data yang terisolasi dari organisasi yang sama.

BAB 8
SISTEM WAJAH ILPB-SVM

Keamanan biometrik telah lama menjadi tren yang memenuhi kebutuhan akan tingkat keamanan dan kontrol yang signifikan. Di antara semua teknologi yang ada, deteksi wajah adalah salah satu inovasi yang paling banyak digunakan dan disesuaikan. Kegagalan identifikasi identitas pengguna menjadi perhatian besar. Dalam bab ini, pendekatan baru untuk pengenalan biometrik telah diperkenalkan di mana penerapan ILBP (*Improved Local Binary Pattern*) untuk deteksi fitur wajah dibahas yang menghasilkan fitur yang lebih baik untuk pola wajah. Ini memungkinkan hanya pengguna yang diautentikasi untuk mengakses sistem, yang lebih baik daripada algoritma sebelumnya.

Penelitian sebelumnya untuk deteksi wajah menunjukkan banyak kekurangan dalam hal tingkat penerimaan dan penolakan yang salah. Dalam makalah ini, ekstraksi fitur wajah dilakukan dari frame statis dan dinamis menggunakan algoritma kaskade Haar. Kemudian, metode ILBP yang bekerja pada nilai piksel lokal suatu gambar diterapkan untuk ekstraksi fitur, dan akhirnya, SVM (*support vector machine*) digunakan untuk klasifikasi fitur tersebut. Tujuan dari makalah ini adalah untuk memberikan hasil pengenalan terbaik dari gambar yang diambil secara acak dan mungkin memiliki noise. Makalah ini mencapai akurasi 97,90% untuk pengenalan yang benar dan dengan kompleksitas waktu yang lebih sedikit. Makalah ini dapat digunakan dalam investigasi kejahatan, kamera keamanan, forensik digital, dll.

8.1 PENDAHULUAN

Fitur biometrik wajah berbasis autentikasi sistem paling umum digunakan di banyak perangkat IoT waktu nyata. Menghadapi banyak ancaman yang terkait dengan penipuan dokumen, ancaman identitas, kejahatan dunia maya, terorisme, dan banyak lagi, teknologi baru telah diterapkan. Di antaranya adalah biometrik, yang digunakan dalam mengautentikasi dan mengidentifikasi seseorang menggunakan karakteristik biometriknya seperti sidik jari, pengenalan wajah, deteksi iris, dll. Fitur subjek biometrik sangat penting untuk pengenalan yang benar. Identitas biometrik adalah identitas unik dan permanen pengguna dan kegagalan pengenalannya bergantung pada banyak keadaan seperti kumpulan data, algoritma, pembelajaran fitur, fitur bajakan, dll. Pengenalan subjek biometrik yang benar dan konsisten adalah tugas yang menantang.

Semua penelitian yang ada tentang pengenalan biometrik memerlukan beberapa titik kunci unik yang cukup menentukan otentikasi yang tepat dari seorang individu. Data biometrik seperti sidik jari, pemindaian iris, sidik telapak tangan, dan wajah subjek umumnya digunakan untuk penelitian otentikasi dan dipercaya secara luas karena mengandung fitur unik. Berbagai gambar biometrik mungkin mengandung noise dan sistem mungkin gagal mengekstraksi fitur titik kunci yang tepat dari gambar biometrik berkualitas rendah yang menyebabkan kegagalan pengenalannya. Karya ini mendukung konsep autentikasi biometrik menggunakan fitur wajah

biometrik. Karya ini tidak hanya menyediakan keamanan autentikasi tetapi juga memenuhi konsistensi dalam pengenalan di bawah berbagai kumpulan data.

Kegagalan pengenalan fitur-fitur tersebut tampaknya terjadi dalam dua kasus yang luas. Pertama, tingkat penerimaan palsu di mana pengguna yang tidak berwenang mendapatkan akses dan kedua, tingkat penolakan palsu di mana pengguna yang berwenang mungkin gagal mendapatkan akses. Kekhawatiran tersebut dalam sistem keamanan sebagian besar dianggap sebagai kesalahan pengenalan yang disebabkan oleh deteksi gambar yang tidak tepat, analisis gambar berkualitas rendah, informasi piksel berkualitas rendah, dll. Sistem pengenalan wajah dilihat di bawah berbagai algoritma pengklasifikasi seperti KNN, AdaBoost, Haar cascading, dll.

Semua model yang didefinisikan sebelumnya tersebut bekerja dengan baik hanya untuk kumpulan data tertentu. Keragaman dalam gambar wajah mungkin tidak mudah dianalisis oleh model dan berakhir dengan kesalahan. Keragaman tersebut dalam kumpulan data dapat berisi gambar piksel berkualitas rendah, gambar yang terganggu, gambar yang difilter, dan variasi acak yang tidak diinginkan lainnya. Untuk mendominasi keragaman tersebut, teknik yang diusulkan menggunakan konsep metode histogram pola biner lokal yang ditingkatkan (ILBP) di mana semua gangguan yang tidak diinginkan pada tangkapan wajah dihilangkan dengan bantuan penguatan atau penekanan fitur berdasarkan nilai intensitas biner.

Teknik yang kuat untuk klasifikasi fitur wajah yang ditingkatkan dirasionalisasi dengan menggunakan algoritma *support vector machine* (SVM) di mana fitur wajah yang diperkuat dari suatu gambar dianalisis berdasarkan nilai geometrisnya dan bidang hiper diputuskan oleh SVM untuk mengklasifikasikan fitur-fitur yang dapat dibedakan. Keputusan tentang kepemilikan fitur-fitur pengujian yang benar diambil oleh SVM yang didasarkan pada fitur-fitur pelatihan. Kombinasi teknik yang kaku ini tidak diragukan lagi nyaman dan kurang rumit untuk pengenalan variasi yang tidak diinginkan dalam gambar wajah. Model yang diusulkan juga berhasil mengenali beberapa subjek wajah secara khas, yang ditangkap dalam satu bingkai.

Biometrik

Biometrik adalah ilmu yang mempelajari penggunaan karakteristik fisik dan perilaku seseorang untuk mengautentikasi dan mengidentifikasi seseorang. Autentikasi biometrik didefinisikan sebagai membandingkan data yang disimpan dalam basis data sebagai templat biometrik dengan karakteristik seseorang untuk menemukan kemiripannya. Basis data disiapkan dengan menyimpan catatan templat biometrik individu yang terdaftar.

- Kemudian data yang disimpan dibandingkan dengan data biometrik individu untuk autentikasi identitasnya.
- Identifikasi biometrik adalah proses mengetahui identitas seseorang.
- Di sini, data biometrik seseorang diambil, yang dapat berupa karakteristik biometrik apa pun.
- Data yang disimpan ini kemudian dibandingkan dengan data biometrik orang lain yang disimpan dalam basis data. Jika ada data yang cocok, pengguna teridentifikasi; jika tidak, seseorang bukan pengguna yang diautentikasi.

Kategori Biometrik

Ada dua kategori biometrik:

- Pengukuran fisiologis

  Pengukuran fisiologis dapat berupa biologis atau morfologis. Pengukuran ini meliputi bentuk morfologis wajah, sidik jari, telapak tangan, mata (retina dan iris), pola vena, dll. Analisis biologis dapat dilakukan melalui air liur, DNA, urin, oleh tim medis dan forensik polisi.
- Pengukuran perilaku

  Pengukuran perilaku umum yang dapat digunakan adalah pengenalan suara, gerakan tubuh, gaya berjalan, dinamika penekanan tombol, dll. Berbagai teknik yang digunakan untuk pengukuran ini merupakan topik penelitian yang sedang berlangsung.

Pengukuran fisiologis biasanya menawarkan manfaat untuk tetap lebih stabil sepanjang hidup seseorang.Pekerjaan ini dilakukan pada fitur wajah biometrik untuk memverifikasi identitas seseorang. Pengenalan wajah adalah salah satu teknik yang paling banyak digunakan dalam analisis gambar. Dalam hal ini, fitur wajah seseorang diekstraksi dan kemudian dikorelasikan dengan gambar yang disimpan dalam basis data sistem. Wajah dideteksi dari gambar dan video. Setelah dideteksi dari gambar, fitur diposisikan dan kemudian cetakan wajah diambil dari fitur wajah. Akhirnya, menggunakan metode klasifikasi objek, data yang diekstraksi dibandingkan dengan data yang disimpan dalam basis data dan metode pengenalan wajah selesai.

Keuntungan Biometrik

- Bersifat universal karena dapat ditemukan pada setiap individu.
- Bersifat unik karena setiap individu memiliki karakteristiknya sendiri.
- Merupakan karakteristik permanen.
- Dapat direkam dalam basis data apa pun.
- Data yang dapat diukur.
- Anti pemalsuan tidak seperti tanda tangan seseorang.

Untuk memungkinkan hanya pengguna yang diautentikasi untuk mengakses data rahasia, teknologi biometrik digunakan dalam banyak perangkat elektronik, BFSI, identifikasi kriminal, bank, keamanan rumah, pertahanan, dll. Teknologi ini bahkan digunakan oleh banyak kantor publik dan swasta dalam bentuk perangkat seperti sidik jari, pengenalan suara, pengenalan wajah, dll. Makna biometrik meningkat dari hari ke hari karena kebutuhan akan keamanan menjadi perhatian penting. Ambil contoh Pemerintah Maharashtra, yang telah membuat basis data penjahat dengan sidik jari, pemindaian retina, dan pengenalan wajah mereka, yang membantu polisi menangkap mereka tanpa penundaan dalam penyelidikan. Ada banyak contoh penggunaan biometrik oleh pemerintah. Satu contoh lagi adalah kartu Adhar, yang menyimpan data individu seperti sidik jari, pemindaian retina, dll., dalam basis data.

Pengenalan Wajah Biometrik

Model pengenalan wajah tidaklah mudah karena kompleksitas dan multidimensinya. Kendala komputasi selalu ada dalam model pengenalan wajah. Pengenalan wajah sama seperti

model pengenalan pola lainnya, tetapi pengenalan wajah berkaitan dengan fitur biometrik. Fokus model selalu pada fitur detail subjek wajah untuk membedakan satu wajah dari yang lain. Tujuan pengenalan wajah adalah identifikasi vektor fitur unik yang memberikan akurasi. Penanganan informasi wajah visual yang tinggi merupakan tugas yang cukup menantang dalam pengenalan wajah.

Informasi visual yang tinggi ini dapat diperoleh dari rekaman video langsung yang menangkap gambar berkualitas tinggi yang membutuhkan waktu komputasi lebih lama untuk menjalankan pengenalan. Pengenalan wajah mengambil serangkaian karakteristik unik yang berada dalam nilai eigen subjek. Nilai-nilai ini dilatih dalam model dan fitur-fiturnya diklasifikasikan dalam kelas-kelas unik. Fitur eigen ini dianggap sebagai serangkaian nilai stabil yang tidak efektif terhadap modifikasi apa pun. Jadi, nilai-nilai ini membantu pengenalan wajah. Modifikasi dapat diperkenalkan pada subjek wajah dalam bentuk serangan pemrosesan gambar atau efek iluminasi apa pun. Namun, nilai eigen ini mungkin dapat menoleransi variasi tersebut dan tidak berubah. Oleh karena itu, nilai-nilai ini mudah digunakan dalam pengenalan wajah. Sistem pengenalan melibatkan pelokalan mata, hidung, mulut, garis luar wajah, dll. Subjek-subjek ini membantu menghubungkan seluruh pengenalan pola wajah oleh model. Ciri-ciri ini sesuai untuk menetapkan hubungan di antara berbagai parameter pola wajah guna mencapai pengenalan yang benar.

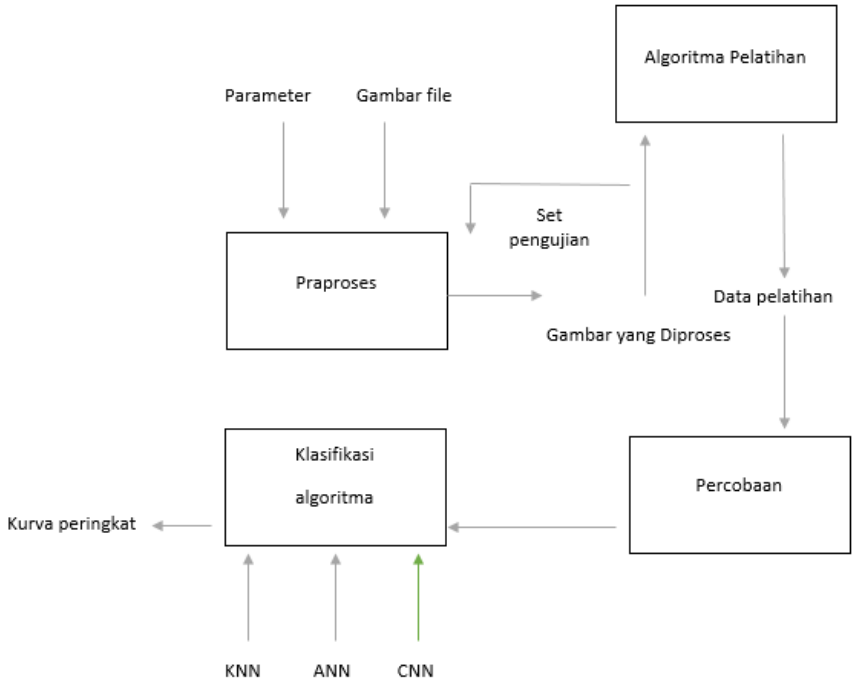
Teknologi yang dipublikasikan sebelumnya menggunakan beberapa strategi pengenalan otomatis dan semi-otomatis berdasarkan metrik jarak yang dinormalisasi di antara titik-titik fitur. Kesulitan dari beberapa tampilan juga diatasi oleh teknik-teknik sebelumnya tersebut. Tujuan pengenalan dicapai dengan menemukan hubungan yang memadai di antara berbagai ciri dan subjek bagian wajah. Pelatihan fitur merupakan salah satu tugas penting di mana fitur dipelajari oleh seluruh jaringan model. Kohonen dkk. membahas sistem pelatihan yang mempelajari unit-unit fitur non-linier dan juga menggunakan algoritma back-propagation untuk memperbaiki kesalahan apa pun dalam pembelajaran fitur. Identifikasi wajah beserta ekspresi juga dibahas oleh Stonham dkk. Penelitian lain tentang pengenalan wajah menggunakan pencocokan templat multi-resolusi yang menggunakan algoritma penginderaan cerdas dan diberikan oleh Burt dkk. Jenis pengenalan ini memenuhi identifikasi dan pemrosesan waktu nyata. Pengenalan wajah didasarkan pada pencocokan fitur uji dengan fitur terlatih dan kemudian pengenalan fitur uji ditetapkan di kelas masing-masing oleh pengklasifikasi. Pencocokan grafik adalah salah satu pendekatan efisien di mana struktur tautan dinamis dibuat berdasarkan pemetaan fitur. Pencocokan tersebut dapat dilihat di pengklasifikasi ANN yang jaringannya dibentuk di antara node.

Pencocokan fitur juga dilakukan menggunakan metrik jarak geometris di mana jarak antara dua nilai piksel dihitung menggunakan jarak Euclidean dan pencocokan dikatakan berhasil hanya jika jaraknya kecil jika tidak, maka akan ditolak. Pencocokan juga dilakukan menggunakan intensitas nilai piksel fitur gambar. Kedua gambar dikatakan sama jika intensitas piksel fitur persilangannya lebih dekat satu sama lain. Kriteria pemetaan berbeda untuk setiap model.

Pengenalan wajah digunakan dalam autentikasi, investigasi kriminal, pemrosesan film, forensik gambar, dll. Jadi, tujuan kami adalah mengembangkan model yang mampu mengidentifikasi setiap wajah secara unik di tengah keramaian dan dalam variasi apa pun. Pendekatan berbasis fitur eigen sangat umum diadopsi oleh peneliti seperti Pentland et al. Nilai eigen ini membantu dalam mengkonstruksi subjek wajah. Tidak ada variasi dan efek yang tidak diinginkan yang disebabkan oleh pencahayaan yang buruk atau posisi wajah orang yang tidak simetris. Visual subjek wajah dicocokkan dengan berbagai cara. Salah satu cara tersebut adalah pencocokan berbasis probabilitas di mana intensitas nilai piksel dipetakan menggunakan algoritma Bayesian.

Algoritma ini digunakan untuk memprediksi pencocokan berdasarkan nilai yang telah ditetapkan sebelumnya. Algoritma ini juga menangani variasi pada gambar yang disebabkan oleh beberapa jenis noise. Jenis algoritma ini menggunakan fungsi kepadatan probabilitas untuk setiap kelas individu dan juga menggunakan versi optimal dari nilai eigen yang sangat efektif dalam pencocokan. Lebih jauh, pengenalan wajah pada dasarnya dilakukan dengan menggunakan pola biner lokal atau fitur LBP yang berisi informasi tekstur gambar. Fitur LBP dari gambar wajah ini diformulasikan dengan bantuan nilai piksel tetangga dari piksel tengah.

Fitur LBP mudah diplot dengan representasi histogram yang menggambarkan informasi tekstur gambar. Skenario pengenalan wajah berisi tiga hal utama, yaitu deteksi wajah, normalisasi, dan identifikasi wajah. Bagian pertama, deteksi wajah di bagian wajah, dideteksi dari seluruh gambar dan disegmentasi. Gambar seseorang berisi gambar latar belakang, rambut, pakaian, dll. Bagian yang menarik hanya wajah. Jadi, pendeteksian bagian wajah dari keseluruhan gambar sangat penting dan juga berfungsi dalam pendeteksian wajah di kerumunan. Pendeteksian dilakukan oleh beberapa algoritma seperti Haar cascade, Viola-Jones, dll. Setelah proses pendeteksian, normalisasi gambar dilakukan di mana piksel gambar dinormalisasi sehingga kita dapat menghindari variasi dan noise yang tidak diinginkan. Langkah terakhir menuju pengenalan wajah adalah pengenalan itu sendiri di mana fitur yang dinormalisasi dilatih dan diklasifikasikan dalam beberapa kelas dan karenanya fitur pengujian dikenali berdasarkan kelas yang dijelaskan oleh pengklasifikasi. Alur dasar pengenalan pola wajah juga dijelaskan oleh diagram alir yang dijelaskan pada Gambar 8.1 yang membantu kita memahami cara kerjanya.



Gambar 8.1 Bagan alir proses pengenalan wajah secara umum.

Dengan diagram di atas, terlihat jelas bagaimana keseluruhan proses pengenalan wajah akan bekerja. Pertama, berkas gambar yang diambil dan parameter terkait seperti ukuran dan resolusinya akan diteruskan ke unit praproses. Fungsi praproses adalah untuk menghilangkan gangguan yang tidak diinginkan dan menghilangkan perbedaan antara gambar. Pada dasarnya, ia mempertahankan simetri gambar wajah. Ia juga melakukan segmentasi jika perlu. Gambar wajah yang diambil mungkin memiliki informasi tambahan yang tidak diinginkan seperti gambar latar belakang atau gambar samping lainnya. Jadi, bagian wajah perlu disegmentasi dari keseluruhan gambar yang diberikan.

Kemudian gambar dibagi menjadi set uji dan set latih. Set latih berisi proporsi yang lebih besar dibandingkan dengan set uji. Blok pelatihan digunakan untuk mempelajari fitur-fitur gambar terlatih yang telah diproses sebelumnya. Fitur-fitur pelatihan ini kemudian diteruskan ke blok eksperimen. Jadi, blok eksperimen digunakan untuk mengevaluasi fitur-fitur fitur pelatihan dan meneruskan fitur-fitur ini ke algoritma pengklasifikasi. Algoritma pengklasifikasi menggunakan semua jenis pengklasifikasi seperti KNN, ANN, CNN, dll., yang digunakan untuk mengklasifikasikan setiap fitur yang dilatih ke dalam kelas-kelas uniknya.

Fitur-fitur uji yang telah diproses sebelumnya kemudian diteruskan ke algoritma pengklasifikasi yang digunakan untuk memprediksi kelas-kelas milik subjek uji. Tugas pengklasifikasi tidak hanya mengklasifikasikan fitur-fitur yang dilatih tetapi juga memprediksi pengenalan kasus uji di kelas-kelas yang diberikan. Hasil pengenalan yang benar dan akurat diformulasikan dalam bentuk kurva peringkat yang tidak lain adalah kurva ROC yang menunjukkan jumlah fitur uji yang cocok dengan benar di kelas-kelasnya masing-masing.

Teknik pengenalan wajah

Teknik pengenalan wajah telah banyak dieksplorasi sebelumnya berdasarkan berbagai pendekatan pembelajaran mesin atau pembelajaran mendalam di mana analisis vektor fitur berlangsung. Teknik seperti itu dinamakan sebagai support vector machine, linear discriminative analysis, Laplacian algorithm, evolution pursuit, dll., yang telah mengakses pengenalan wajah dengan sukses. Semua teknik tersebut sedikit rumit dan lebih memakan waktu untuk gambar yang samar atau berkualitas tinggi. Dan mereka tidak menjamin akurasi pengenalan untuk variasi gambar yang beragam. Selain itu, Hallinan et al. mengusulkan teknik pengenalan wajah berdasarkan fitur eigenface yang dilakukan dengan variasi kondisi pencahayaan, tetapi metode ini tidak dapat menjamin pengenalan gambar kabur yang mengandung informasi wajah paling sedikit. Belhumeur et al. juga membahas pengenalan wajah berdasarkan teknik subruang linier 3D yang bekerja dengan baik untuk pencahayaan dan orientasi variabel. Tetapi teknik ini sedikit rumit dan memiliki waktu komputasi yang tinggi karena bekerja pada subruang 3D.

Model berbasis SPD dan manifold Grassmann mempertimbangkan semua gambar dengan kualitas yang sama. Metode ini bagus untuk gambar berkualitas rendah karena meningkatkan semua gambar ke tingkat yang sama. Namun, itu tetap tidak membenarkan pengenalan untuk bingkai yang diambil dengan beberapa wajah. Memuaskan pengenalan data wajah secara khusus di antara kerumunan adalah salah satu masalah utama dalam sistem pengenalan, juga membahas metode yang efektif untuk lokalisasi wajah menggunakan pembelajaran penguatan, tetapi sedikit rumit karena bekerja berdasarkan pohon keputusan dan karenanya menghabiskan lebih banyak waktu untuk melokalisasi subjek wajah.

Teknik yang diusulkan dari makalah ini menggunakan pengklasifikasi cascading Haar untuk melokalisasi bagian wajah dari gambar yang diambil dan teknik ini jauh lebih sederhana karena menggunakan fitur Haar untuk menutupi blok persegi panjang subjek wajah. Pengenalan wajah yang bergantung pada subruang manifold dan sparse atau linear juga merupakan beberapa teknik usang yang gagal menoleransi tingkat kesalahan penolakan palsu dan penerimaan palsu.

Kontribusi Utama

Kontribusi makalah ini tidak hanya terletak pada gambar statis tetapi juga pada bingkai video dinamis waktu nyata. Data dari video dapat ditangkap, dideteksi, dan diklasifikasikan dengan tingkat kesalahan minimum. Gambar wajah diekstraksi menggunakan algoritma kaskade Haar yang menemukan fitur haar pada subjek wajah dari seluruh gambar. Teknik yang diusulkan memperkenalkan kombinasi baru dari desain ILBP dan SVM yang tangguh. Algoritma ILBP menyediakan versi fitur lokal pola wajah yang lebih baik untuk model SVM.

SVM digunakan sebagai pengklasifikasi yang memberikan hasil terbaik dalam hal pengenalan fitur lokal yang lebih baik tersebut dengan kompleksitas waktu yang lebih sedikit. Deteksi wajah, ekstraksi fitur, dan pengenalan dilakukan secara efisien menggunakan algoritma canggih yang mempertahankan akurasi pengenalan pada berbagai noise dalam gambar sambil mempertahankan kompleksitas waktu. Akurasi pengenalan ditemukan sebesar 97,90% yang lebih baik daripada algoritma lain yang ada.

Fitur Terbaru

Versi perbaikan fitur biner lokal diekstrak dari gambar wajah pengguna. Fitur-fitur ini sangat stabil dan tidak menyebabkan distorsi apa pun pada modifikasi apa pun. Fitur-fitur ini tidak mudah dilacak. Konsep di balik penggunaan fitur-fitur ini adalah konsistensinya. Ini adalah fitur-fitur yang disempurnakan dan membantu meningkatkan ketahanan autentikasi model untuk gambar acak juga. Gagasan untuk menggunakan pola ILBP fitur wajah disertai dengan keterbatasan dalam model klasifikasi. Pada model-model sebelumnya, pengenalan dan klasifikasi fitur wajah tampaknya gagal untuk gambar wajah acak yang mungkin mengandung gangguan acak.

Dengan kata sederhana, gambar wajah yang diambil secara acak dalam waktu nyata dari kerumunan mungkin mengandung banyak variasi yang tidak diinginkan seperti efek pencahayaan, pencahayaan latar belakang, pengaburan, orientasi yang tidak tepat, pose yang tidak tepat, malaikat, dll. Dengan kata sederhana lainnya, gambar pengguna yang diklik secara acak yang diklik secara real time dari kerumunan berbeda dari gambar yang diklik biasa yang diklik di latar belakang simetris dengan pencahayaan dan jarak yang konstan. Oleh karena itu, tugas yang menantang adalah mengamankan pengenalan gambar wajah yang diklik secara acak tanpa parameter simetris dan karenanya mengandung variasi acak (Bharat Singh et al., 2013). Pekerjaan yang diusulkan berkontribusi untuk mengamankan pengenalan dan klasifikasi gambar tersebut. Oleh karena itu, model yang diusulkan menggunakan fitur ILBP yang diekstraksi dari gambar wajah acak tersebut, dan dengan menggunakan SVM, pekerjaan klasifikasi berlangsung. Ide di balik penggunaan SVM adalah untuk menghemat waktu komputasi karena gambar wajah tersebut tidak didefinisikan sebelumnya.

Prediksi waktu pemrosesannya tidak selalu sama karena tergantung pada variasi yang dimuat pada gambar saat diambil dengan kamera. SVM adalah program yang ringan dan mampu menghasilkan bidang hiper dengan mudah untuk sejumlah besar vektor fitur. Meskipun teknik sebelumnya mampu menghasilkan tingkat akurasi pengenalan yang cukup baik, model yang diusulkan memberikan tingkat akurasi pengenalan sebesar 97,90% yang merupakan kriteria yang dapat diterima dan juga bersaing dengan beberapa penelitian terbaru tahun 2018, 2019, dan 2020 dalam model pengenalan wajah.

Kebaruan model yang diusulkan tidak terbatas pada tingkat akurasi tetapi meluas hingga ke kumpulan data juga. Pekerjaan yang diusulkan menggunakan kumpulan data buatan sendiri yang berisi gambar acak yang diklik dari orang yang berbeda, seperti yang dibahas sebelumnya. Gambar-gambar ini tidak diklik dengan jarak tertentu atau dalam kondisi lingkungan tertentu. Oleh karena itu, model yang diusulkan mengamankan konsistensi pengenalan gambar-gambar tersebut yang berisi variasi acak (seperti dalam rekaman waktu nyata) dengan akurasi 97,90%. Perbedaan utama antara teknik lain dan model yang kami usulkan bukan hanya tingkat akurasi tetapi juga jenis kumpulan data dan fitur yang diambil.

Dalam teknik lain yang ada, penelitian berada dalam kumpulan data simetris yang berisi gambar simetris yang diklik dalam kondisi lingkungan tertentu. Konsistensi pengenalan oleh model yang ada mungkin tidak cukup kuat untuk memproses gambar-gambar yang berisi variasi acak yang besar. Stabilitas pengenalan semua model pembelajaran mesin didasarkan

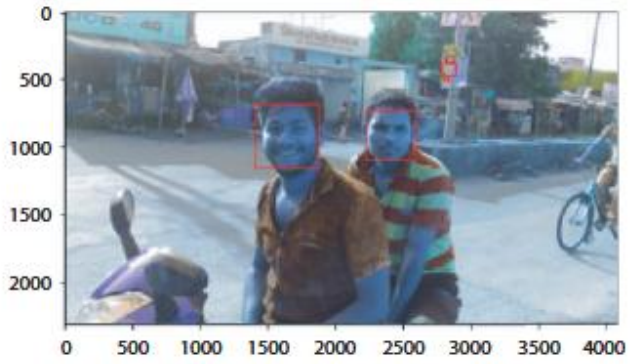
pada fitur pembelajaran atau pelatihannya. Model-model yang ada tersebut hanya dapat berfungsi jika pelatihan yang diperlukan diberikan kepada mereka. Namun, ini merupakan beban tambahan untuk melatih model pembelajaran mesin berulang-ulang untuk kumpulan data variabel. Untuk menghilangkannya, model yang diusulkan menggunakan fitur ILBP yang merupakan fitur yang paling stabil dan halus yang diekstraksi untuk gambar yang tidak simetris, dan fitur tersebut tidak diharapkan untuk dimodifikasi di bawah gangguan.

Kebaruan lain dari pekerjaan yang diusulkan adalah mendeteksi bagian wajah dari gambar yang diklik. Gambar yang diklik juga memuat bagian tubuh orang lain dan objek latar belakang. Karena karya yang diusulkan menggunakan gambar acak yang diklik secara non-simetris, gambar seseorang dapat membawa objek samping lain atau orang lain di sekitarnya. Setiap gambar acak yang diklik dari jalan secara real time memuat objek, orang lain, kendaraan, latar belakang gelap, efek pencahayaan, dll.

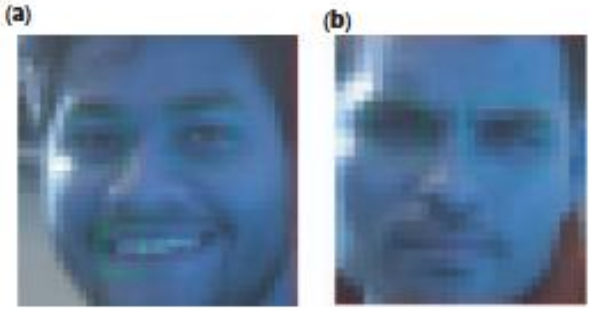
Oleh karena itu, tugas menantang lainnya adalah mendeteksi dan mengelompokkan bagian wajah yang diperlukan dari seseorang dari seluruh gambar yang diklik. Model yang diusulkan menggunakan pengklasifikasi Haar-cascade untuk mendeteksi bagian wajah dari gambar secara khas. Teknik ini pertama-tama memplot fitur Haar dari bagian wajah dan dengan mendeteksi bentuk hidung dan mata, ia menandai batas persegi panjang di sekitar bagian wajah. Oleh karena itu, dalam kasus kerumunan, setiap detail wajah dapat dengan mudah disegmentasi tanpa kesalahan ketidakcocokan. Dalam penelitian yang ada, tugas deteksi tidaklah menantang karena gambar diambil dalam kondisi lingkungan tertentu (Jyh-Yeong Chang et al.) (Pramod Singh Rathore et al., 2017). Makalah yang tersisa dibagi menjadi berikut: Bagian 2 mendefinisikan metodologi yang diusulkan, bagian 3 berisi data eksperimen, bagian 4 membahas kesimpulan, yang diikuti oleh referensi.

8.2 DETEKSI WAJAH MENGGUNAKAN ALGORITMA HAAR

Deteksi subjek wajah dari seluruh gambar dilakukan oleh algoritma kaskade Haar. Deteksi subjek wajah bergantung pada fitur gambar yang digunakan algoritma ini. Pengklasifikasi ini diberikan oleh Viola dan Michael Jones. Nilai Fitur dihitung oleh pengklasifikasi Haar menggunakan integral persegi panjang yang mengalikan bobot setiap persegi panjang dengan nilai luasnya dan menjumlahkan semuanya. Algoritma bekerja dengan mendeteksi beberapa fitur penting wajah dan menggambar persegi panjang di sekitarnya. Algoritma dilatih dengan beberapa gambar positif dan negatif dan kemudian bergantung pada gambar tersebut, fitur dideteksi. Gambar 8.2 dan 8.3 menunjukkan subjek wajah yang terdeteksi dan tersegmentasi dari gambar input.



Gambar 8.2 Mendeteksi fitur wajah dari sebuah gambar.

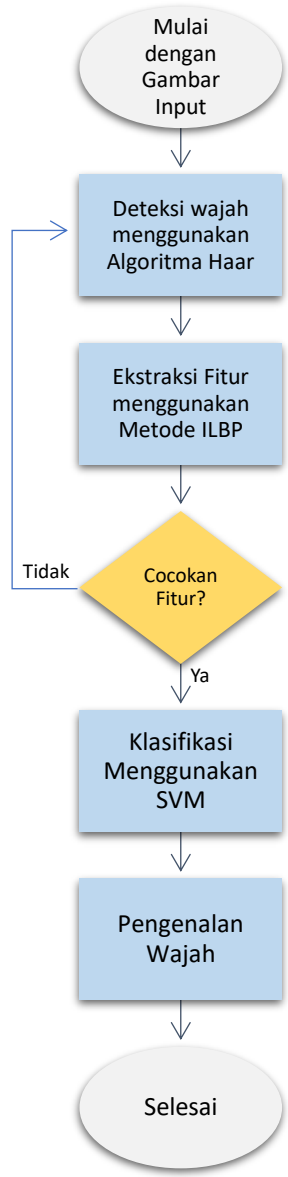


Gambar 8.3 Mengekstrak gambar yang terdeteksi.

Fitur Haar diekstraksi dengan mengelilingi piksel yang diperlukan dengan struktur persegi panjang. Struktur persegi panjang mencakup semua piksel yang dibutuhkan di empat sudut dan menjumlahkan semuanya. Proses perhitungan dapat dilihat dari persamaan yang diberikan di bawah ini.

$$\sum_{a=j0}^{j1} W \sum_{b=k0}^{k1} I\,(a,b) = F(j1,k1) - F(j0,k1) - F(j0,k1) + F\,(j0,k0$$

Dimana I merupakan bayangan utuh dan jumlah intensitasnya berkisar dari $(x0, y0)$ sampai $(x1, y1)$ yang dapat dilihat pada Gambar 8.4.



Gambar 8.4 Diagram alir Sistem.

Gambar 8.5 Integral persegi panjang.

Keempat nilai persegi panjang dijumlahkan untuk mendapatkan fitur Haar tertentu yang ditunjukkan pada Gambar 8.5.

Ekstraksi Fitur Menggunakan ILBP

Makalah ini mengusulkan metode baru untuk mengekstrak fitur batas gambar. Tidak diinginkan untuk mendapatkan matriks gambar berdimensi tinggi. Hanya fitur batas penting dari suatu objek yang perlu diekstraksi. Karakteristik batas diperoleh dengan menggunakan ILBP yang memiliki dimensionalitas rendah. Gambar yang dipilih untuk percobaan mungkin mengalami variasi yang beragam terhadap efek iluminasi dan beberapa noise tertentu merupakan hasil dari ukuran, translasi, rotasi, dan ekspresi acak. Faktor-faktor ini dapat membatasi pengenalan fitur batas suatu gambar. Model yang diusulkan memperkenalkan algoritma ILBP yang mengekstrak definisi lokal titik-titik kunci wajah di bawah gangguan variabel. Teknik Pola Biner Lokal yang Disempurnakan berakar pada penentuan pola 2D. Dalam ILBP, nilai rata-rata piksel dibandingkan dengan nilai piksel di dekatnya atau di lingkungan sekitar untuk mendapatkan perbedaan paling kecil dengan semua nilai piksel lainnya. Ini membantu dalam mendapatkan jarak nilai piksel rata-rata dengan semua nilai piksel lainnya yang menghasilkan pola yang lebih baik. Dalam ILBP, seluruh piksel lingkungan 3 x 3 diberi ambang batas dengan nilai skala abu-abu rata-rata dan menyediakan 29-1 kemungkinan pola. Pekerjaan proses ini didefinisikan dalam persamaan 2.

$$(X) = \begin{cases} 1, & x \geq 0 \\ 0, & x \geq 0 \end{cases}$$

$$\mathrm{f_{ILBP}(X)} = 2^8 \xi(\mathrm{lc} - \mathrm{S}) + \sum_{\mathrm{J}=0}^{7} 2\mathrm{j}\ \xi(\mathrm{Ij} - \mathrm{S}) - 1$$

di mana Ic adalah nilai piksel tengah, In adalah nilai piksel lingkungan, dan S adalah nilai skala abu-abu rata-rata yang dihitung dalam persamaan 3.

$$S = 1/9\ lc(+\sum_{j=0}^{7} Ij)$$

ILBP merepresentasikan vektor fitur dalam bentuk pola biner 0 dan 1. Pola biner ini diubah menjadi bentuk desimal dan disimpan dalam matriks yang merepresentasikan vektor fitur, seperti yang ditunjukkan di bawah ini.

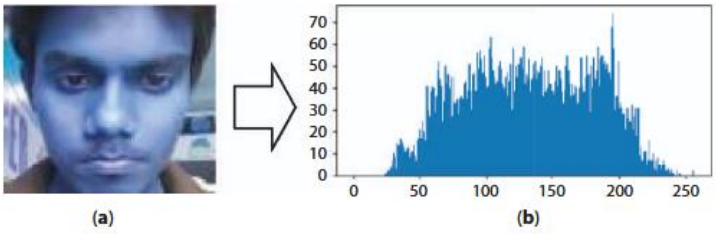
205	200	90
174	111	89
80	236	112

Gambar Skala Abu-abu Nilai piksel Skala Abu-abu

$$(205 + 200 + 90 + 89 + 112 + 236 + 80 + 174 + 111)/9 = 149{,}67$$
$$(111 - 149{,}67) = \text{nilai negatif} = 0$$
$$(112 - 149{,}67) = \text{nilai negatif} = 0$$
$$(205 - 149{,}67) = \text{nilai positif} = 1$$
$$(174 - 149{,}67) = \text{nilai positif} = 1$$
$$(89 - 149{,}67) = \text{nilai negatif} = 0$$
$$(80 - 149{,}67) = \text{nilai negatif} = 0$$
$$(90 - 149{,}67) = \text{nilai negatif} = 0$$
$$(236 - 149{,}67) = \text{nilai positif} = 1$$
$$(200 - 149{,}67) = \text{nilai positif} = 1$$

$$(010100011)_2 = 163_{10} \rightarrow \text{mewakili Pola}$$

Dari citra wajah yang diekstrak, vektor fitur masing-masing diekstrak dalam bentuk pola biner dan setelah mengubahnya menjadi pola desimal, vektor tersebut disimpan dalam matriks. Fitur-fitur tersebut direpresentasikan oleh histogram seperti yang diberikan pada Gambar 8.6.



Gambar 8.6 (a) Citra tersegmentasi (b) Histogram fitur yang dihasilkan oleh ILBP.

Kumpulan data

Teknik yang diusulkan diuji pada kumpulan data kami sendiri yang kami buat menggunakan klik acak. Citra-citra ini tidak memiliki simetri apa pun dan mengandung derau di dalamnya. Basis data terdiri dari 200 citra. Kami memiliki 10 kelas yang masing-masing memiliki 20 citra. Untuk menguji ketahanan teknik kami, kami juga mempertimbangkan citra dari bingkai video. Semua citra disimpan dalam satu direktori. Citra basis data diubah menjadi skala abu-abu dan kemudian proses ekstraksi fitur dilakukan. Beberapa citra dari kumpulan data didaftarkan dalam model untuk tujuan pelatihan guna mempelajari fitur-fiturnya dan

kemudian citra acak lainnya digunakan untuk tujuan pengujian. Jika citra yang diberikan ditemukan dari citra yang dilatih, maka wajah akan dikenali; jika tidak, citra akan ditampilkan sebagai orang yang tidak dikenal. Citra contoh basis data ditunjukkan pada Gambar 8.7.

Klasifikasi Menggunakan SVM

makalah ini, kami menggunakan SVM untuk melatih model kami, yang memberikan akurasi yang lebih baik. Kami mempertimbangkan 200 gambar yang dibagi menjadi total 10 kelas yang masing-masing memiliki 20 gambar. Gambar-gambar ini diambil dari klik acak dan beberapa di antaranya diambil dari bingkai video juga. Gambar yang diambil memiliki variasi, ekspresi, pencahayaan yang berbeda dan juga memiliki noise yang disebabkan oleh faktor eksternal. Setelah ekstraksi fitur, klasifikasi dilakukan oleh SVM (*support vector machine*). SVM pada dasarnya membentuk hyper-plane yang memisahkan fitur-fitur diskriminan. Asumsikan, (X_i, Y_i) adalah kumpulan sampel fitur di

Gambar 8. 7 Gambar kumpulan data orang.

mana $Y = (^{+1}_{-1})$ adalah label kelas. Persamaan hyper-plane diberikan di bawah ini:

$$(W \,.\, X) + b = 0$$

Fungsi klasifikasi optimal dari algoritma SVM diberikan di bawah ini.

$$g(x) = sng\left(\sum_{i}^{n} Y_i a_i^0 \,(X_i \,.\, X) - b_0\right)$$

Di sini, X adalah vektor pendukung atau SV_s, a_i^0 adalah koefisien bahasa yang sesuai, b_0 adalah nilai ambang batas. Persamaan ini menetapkan kasus yang dapat dipisahkan secara linier untuk setiap fitur diskriminan. Dalam klasifikasi SVM, ruang berdimensi tinggi dipetakan dengan vektor input dengan bantuan transformasi non-linier. Dengan cara ini, hiperbidang yang optimal dapat diperoleh. Jika fungsi produk dalam atau fungsi kernel digunakan sebagai ganti penggunaan produk titik dalam klasifikasi optimal, maka fungsi diskriminan yang sesuai akan menghasilkan yang berikut:

$$g(x) = sng\left(\sum_{i}^{n} Y_i a_i^0 \,K(X_i \,.\, X) - b)\right)$$

Ini menunjukkan fungsi diskriminatif dari *Support Vector Machine*. Dimungkinkan untuk menghasilkan berbagai fungsi SVM diskriminatif nonlinier menggunakan fungsi kernel yang berbeda $K(X_i, X)$. Fungsi yang dihasilkan berikut diberikan di bawah ini:
Fungsi polinomial:

$$K(X_i, X) = [(X, X_i) + 1]^d$$

Di sini, SVM adalah pengklasifikasi polinomial derajat d. Fungsi radial:

$$K(X_i\, X) = \exp\{-|X - X_i|^2/a^2$$

Di sini, SVM bertindak sebagai pengklasifikasi Gaussian RBF. Fungsi sigmoid:

$$K(X_{i,}\, X) = \tanh(v(X_i.\, X) + c)$$

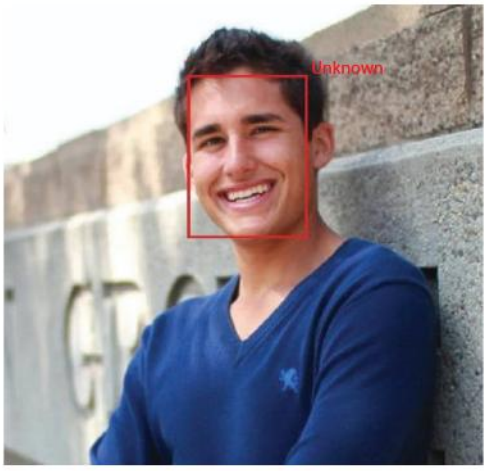
Di sini, SVM bertindak sebagai multi-layer perceptron. $X_i = (X_1, X_2, X_3 \ldots X_m)$ adalah vektor pendukung. $K(X_{i,}\, X)$ menunjukkan fungsi kernel. Hasil dari SVM adalah:

$$Y = \text{sgn}\left(\sum_{i=1}^{n} Y_i\, a_i K(X_{i,} X) - b\right)$$

Dengan menggunakan metrik jarak, kesamaan antara fitur yang diuji dan fitur yang dilatih dihitung dan ditemukan jarak Euclidean antara nilai fitur, kemudian dibandingkan dengan nilai ambang batas. Akurasi ditingkatkan dengan mengubah nilai ambang batas. Jarak yang dihitung harus lebih kecil dari nilai ambang batas untuk menunjukkan kesamaan. Oleh karena itu, berdasarkan kesamaan antara masing-masing kelas, gambar diklasifikasikan ke dalam kelasnya masing-masing. Setiap kelas secara terpisah termasuk dalam identitas individu yang menandakan kepemilikan yang sebenarnya.

8.3 EKSPERIMEN METODE PENGENALAN WAJAH

Kami telah membuat kumpulan data kami sendiri yang terdiri dari 200 gambar dengan 10 kelas individu yang masing-masing memiliki 20 gambar orang yang berbeda. Kumpulan data tersebut juga berisi gambar yang diambil dari bingkai video. Setiap gambar kemudian diberi anotasi dengan label yang berbeda. Kumpulan data dibagi dalam rasio 60:40 untuk set pelatihan dan pengujian. 120 gambar digunakan untuk pelatihan model SVM dan 80 gambar digunakan untuk tujuan pengujian. Dimulai dengan pengumpulan data set dan ekstraksi fitur menggunakan vektor fitur ILPB, kemudian wajah dideteksi dan diklasifikasikan ke dalam kelasnya masing-masing. Kemudian gambar yang diuji diberikan sebagai input dan jika fiturnya cocok dengan fitur gambar yang dilatih, maka gambar tersebut diklasifikasikan sebagai gambar yang dikenal dan kelas yang dimiliki; jika tidak, model mengenalinya sebagai wajah yang tidak dikenal seperti yang ditunjukkan di bawah ini.



Gambar 8.8 Wajah yang tidak dikenal.

Gambar 8.8 menunjukkan gambar yang tidak ada dalam kumpulan data dan diuji sebagai orang yang tidak dikenal. Hasil eksperimen dibagi menjadi tiga bagian berikut:

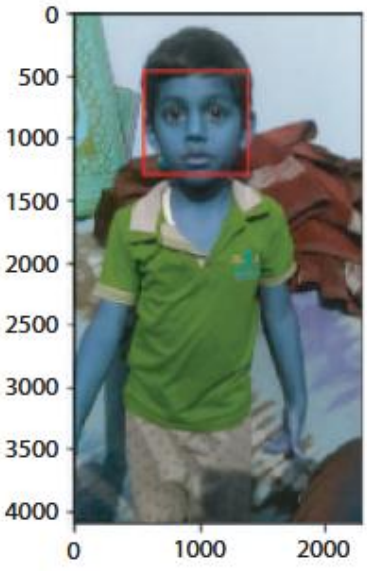
- Deteksi subjek wajah
- Ekstraksi Fitur
- Pengenalan Wajah

Deteksi Wajah

Wajah dari gambar dideteksi dan diekstraksi menggunakan pengklasifikasi kaskade Haar. Contoh gambar yang terdeteksi dan tersegmentasi diberikan pada Gambar 8.9 dan Gambar 8.10, masing-masing.

Ekstraksi Fitur

Pada bagian ini, fitur diekstraksi dari bagian wajah gambar menggunakan metode ILBP, yang disimpan dalam bentuk matriks vektor fitur. Tabel 8.1 menunjukkan contoh gambar yang terdeteksi dan diekstraksi dan masing-masing fitur ILBP yang direpresentasikan oleh histogram.



Gambar 8.9 Deteksi wajah.

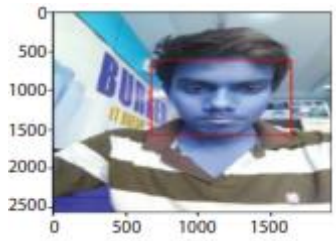
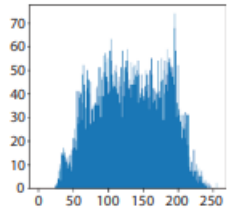
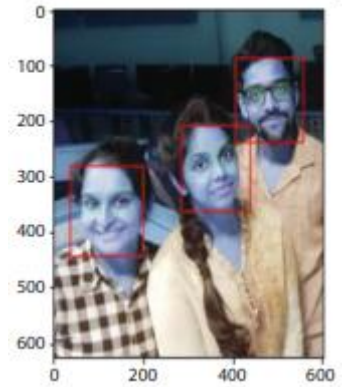
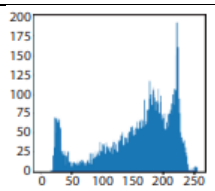
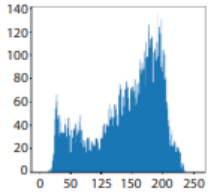
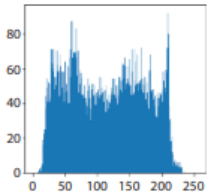
Gambar 8.10 Citra wajah tersegmentasi.

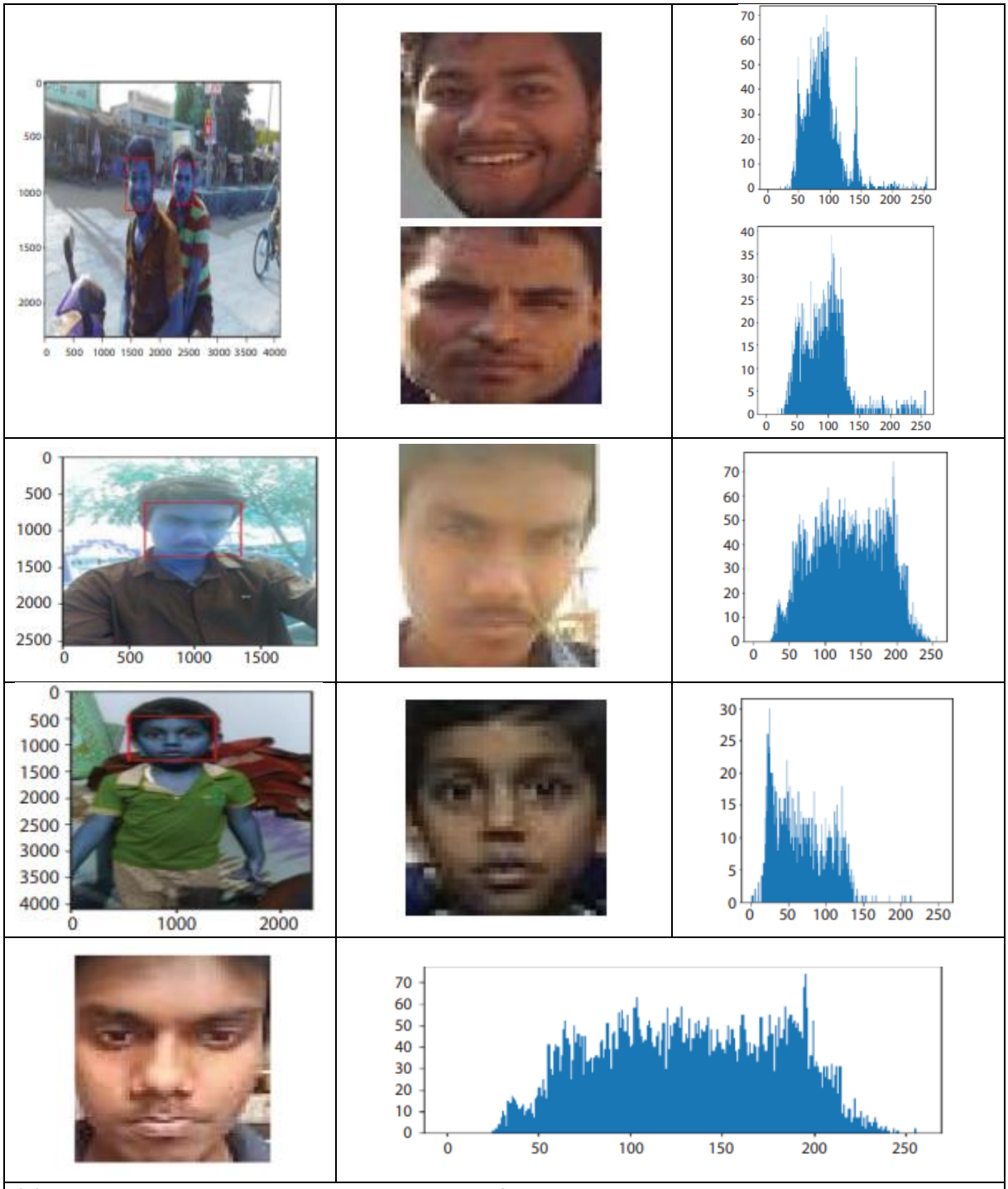
Gambar 8.11 menunjukkan perbandingan antara fitur yang dihasilkan dari dua metode berbeda, yaitu LBP yang ditingkatkan dan LBP sederhana]. Dari Gambar 8.10, disimpulkan bahwa algoritma ILBP menghasilkan fitur yang lebih baik dibandingkan dengan fitur yang diekstraksi oleh algoritma LBP sederhana.

Mengenali Citra Wajah

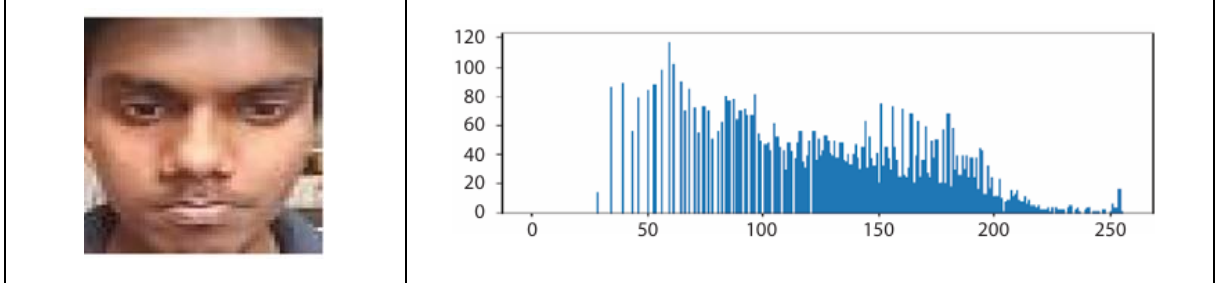
Pengenalan fitur wajah dan kemudian klasifikasinya merupakan hasil akhir dari percobaan. Jika fitur tersebut cocok dengan fitur yang dilatih dalam basis data, maka keluaran akan diperoleh dari basis data sebagai wajah yang dikenali. Kurva ROC yang diberikan pada Gambar 8.12 menunjukkan akurasi pencocokan fitur yang dilakukan oleh algoritma SVM.

Tabel 8.1 Contoh ekstraksi fitur menggunakan metode ILBP.

Contoh gambar dari dataset	Gambar yang diekstraksi (RGB)	Histogram fitur
		
	  	  



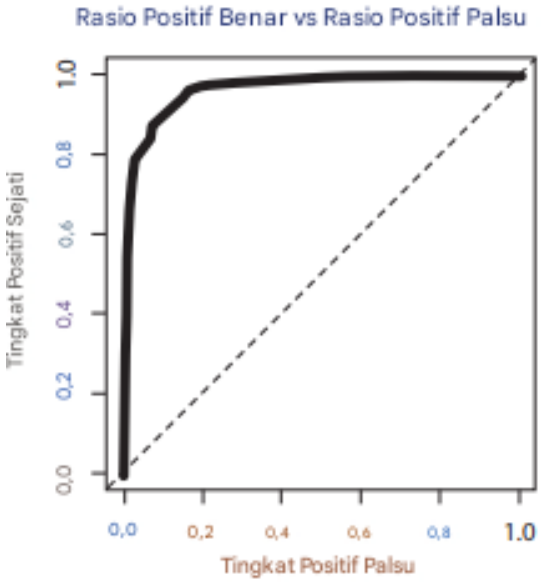
(a) Contoh gambar tersegmentasi dan histogram fiturnya menggunakan algoritma ILBP



(b) contoh gambar tersegmentasi dan histogram fiturnya menggunakan algoritma LBP sederhana

Gambar 8.11 Perbandingan antara fitur yang diekstraksi menggunakan (a) metode ILBP dan (b) metode LBP sederhana.

Menurut Gambar 8.12, akurasi model pada set data uji ditemukan sebesar 97,90%. Beberapa gambar yang diuji terdiri dari data yang tidak diketahui. Tingkat kesalahan dalam pengenalan hanya 2,10%. Oleh karena itu, model yang diusulkan mampu membuktikan akurasi pengenalan yang baik untuk berbagai jenis gambar yang mungkin mengandung noise yang tidak dapat diprediksi.



Gambar 8.12 Kurva ROC untuk pengenalan wajah.

Tabel 8.2 Tabel perbandingan.

No. Urut	Teknik	Akurasi/Hasil
1	Pengenalan wajah berbasis Rutin Kontur Kasar dan Estimasi	92.1%
2	Analisis komponen utama berdasarkan pengenalan wajah	83%
3	AdaBoost dan analisis komponen lanjutan berdasarkan pengenalan wajah	95.50%
4	Transformasi Simetri Radial	83%
5	Pengenalan berbasis SVM	95.71%
6	Teknik yang diusulkan	97.90%

Tabel 8.2 menunjukkan tabel perbandingan di mana teknik yang diusulkan dibandingkan dengan teknik lain yang telah dipublikasikan dan menunjukkan hasil yang baik dibandingkan dengan teknik lainnya.

Dalam bab ini, ILBP dengan SVM digunakan terutama untuk mengidentifikasi fitur wajah seseorang. Eksperimen dilakukan dengan menerapkan ILBP bersama dengan SVM untuk mengklasifikasikan hasil dengan akurasi terbaik. Jadi untuk ini, kami telah menguji pendekatan kami pada beberapa gambar dan kami sampai pada kesimpulan bahwa dengan ILBP fitur dapat

diekstraksi dengan cara yang lebih baik, yang menghasilkan akurasi yang lebih baik sebesar 97,90%. Dan juga algoritma SVM merupakan pengklasifikasi yang kuat yang melatih kumpulan data secara efisien dan mengklasifikasikan data yang diuji secara akurat dan dengan tingkat kesalahan yang lebih rendah. Akurasinya akan meningkat jika kumpulan data pelatihannya ditingkatkan.

Salah satu penggunaan pengenalan wajah yang paling terkenal adalah untuk keamanan. Tenaga kerja pelaksana hukum dapat memanfaatkan inovasi ini untuk mengenali dan membedakan orang dengan memeriksa siapa pun yang memasuki klinik medis. Mereka kemudian akan dapat membandingkan setiap individu dengan ikhtisar orang yang dikenali. Inovasi ini juga dapat digunakan di klinik medis untuk mengenali orang-orang yang mungkin mencari obat terlarang atau orang-orang yang baru saja keluar dari rumah sakit yang tidak lagi diizinkan oleh klinik darurat untuk datang ke kantor. Rumah sakit juga dapat menggunakan inovasi pengenalan wajah untuk membedakan desain yang melibatkan wawasan umum di sekitar tamu dan pasien tergantung pada jenis kelamin dan usia. Kerangka kerja ini dapat memungkinkan kantor untuk mengikuti pasien tanpa menggunakan suar GPS fisik. Ini dapat terbukti berguna untuk menemukan pasien di dalam panti jompo atau di kantor rawat jalan atau panti jompo.

BAB 9
DETEKSI CACAT KABEL DAN PIPA MEDIS BERBASIS ANN DAN IOT

Robot dengan inspeksi penglihatan telah dikembangkan untuk mendeteksi cacat pada kabel multi-untai pada kabel bentang panjang. Sistem yang dikembangkan terdiri dari robot pemanjat, kamera untuk menangkap gambar, modul IoT untuk mengirimkan gambar ke cloud, platform pemrosesan gambar, dan modul jaringan saraf tiruan yang ditujukan untuk pengambilan keputusan. Robot pemanjat memegang kabel dengan roda beralur bersama dengan kamera pemicu otomatis dan modul IoT. Untuk inspeksi, robot naik di sepanjang kabel secara terus-menerus dan memperoleh gambar berbagai segmen kabel. Kemudian gambar yang diambil telah dikirim ke penyimpanan awan melalui sistem IoT. Gambar yang disimpan telah diambil dan ukurannya telah diperkecil melalui teknik pemrosesan gambar.

Data gambar yang diperkecil telah disediakan sebagai respons masukan ke modul jaringan saraf tiruan untuk pengambilan keputusan tentang identifikasi cacat. Hasil eksperimen yang diperoleh menunjukkan dan membuktikan bahwa teknologi inspeksi robot penglihatan cerdas yang diproyeksikan adalah yang paling sesuai untuk inspeksi dan penilaian kondisi kabel multi-untai pra-tekanan. Robot yang dikembangkan juga dapat digunakan untuk memeriksa sistem pipa gas medis. Robot telah diprogram untuk mendeteksi warna pipa O2, N2, vakum, dan tekanan udara.

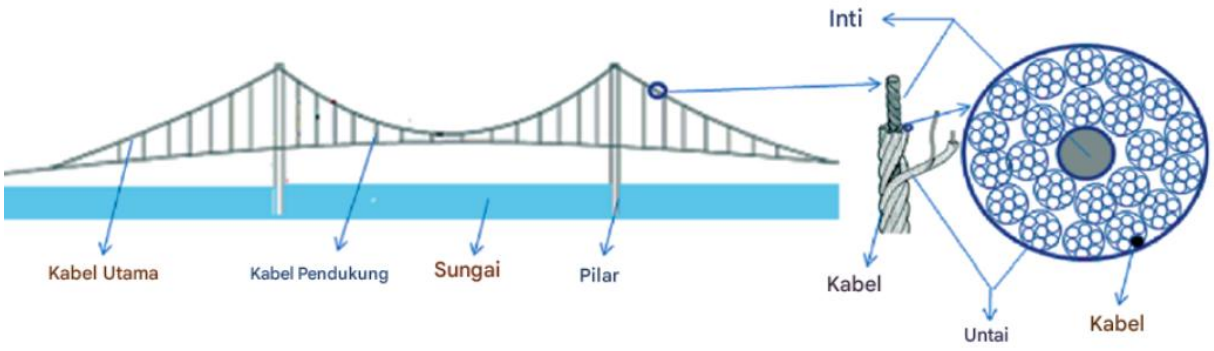
9.1 PENDAHULUAN

Karena globalisasi dan kemajuan dalam industri otomotif, industri transportasi, dan logistik, transportasi jalan menjadi bagian penting dari kehidupan sehari-hari, baik dalam domain domestik maupun industri. Jadi jembatan dibangun di atas sungai dan lembah untuk transportasi cepat dan untuk menghubungkan berbagai segmen negara. Dengan demikian jembatan menjadi bagian penting dari transportasi saat ini untuk menghubungkan semua bagian negara. Volume lalu lintas jalan yang lebih besar di negara-negara berkembang dan lingkungan yang ekstrem menyebabkan kerusakan struktur jembatan. Jadi muncul kebutuhan untuk pemantauan dan pemeliharaan struktural berkala untuk jembatan. Pada jembatan kabel tetap, kabel multi-untai pra-tekanan adalah bagian pembawa beban kritis yang memerlukan pemantauan dan pemeliharaan yang lebih tinggi.

Kabel multi-untai pra-tekanan ini terdiri dari kumpulan untai baja berkekuatan tinggi, yang dipilin satu sama lain untuk menciptakan pratekan permanen pada kabel; dengan demikian kabel tidak akan mengalami deformasi akibat beban besar. Kerusakan pada kabel atau untaian atau bagian apa pun pada kabel akan berkembang lebih cepat dan dengan demikian menjadi ancaman yang lebih besar bagi keselamatan jembatan. Bahan kabel mudah rusak, karena berada dalam kondisi pra-tekanan dan di sektor kelembaban tinggi. Jadi kerusakan kabel harus dideteksi tepat waktu dan penting untuk memahami keadaan saat ini, bersama dengan kemungkinan masa pakai yang akan datang untuk menghindari bencana.

Di beberapa jembatan, pembungkus baja atau penutup PVC telah digunakan untuk menutupi dan mencengkeram untaian kawat dengan kuat. Saat ini, metode pemeriksaan non-destruktif seperti sinar-X, fluks magnetik, sinar gamma, arus eddy, ultrasonik, dll., telah dilakukan untuk mendeteksi kerusakan pada kabel. Meskipun metode tersebut menghasilkan hasil yang memuaskan, terkadang hambatan telah terjadi di bagian dalam pembungkus baja untuk sinyal, yang dapat menyebabkan hasil yang salah. Selain itu, lingkar luar kabel multi-untai terlalu besar dan dengan demikian mencegah metode pemeriksaan non-destruktif.

Dalam beberapa kasus, pipa polietilena digunakan sebagai lapisan pelindung untuk kabel jembatan. Karena penuaan atau kecelakaan, jika ada retakan yang terbentuk di pipa pelindung, zat korosif akan mulai menembus ke dalam kabel multi-untai internal, dan merupakan tugas yang menantang untuk mendeteksi kerusakan. Metode pemindaian laser digunakan untuk mendeteksi kerusakan yang efisien, tetapi sistemnya relatif sangat besar, sulit ditangani, dan lebih mahal. Gambar jembatan kabel ditunjukkan pada Gambar 9.1.



Gambar 9.1 Jembatan kabel.

Dengan demikian, muncul kebutuhan untuk mengembangkan model inspeksi otomatis yang lebih kompeten untuk mendeteksi cacat permukaan pada kabel jembatan dengan biaya rendah dan perawatan yang lebih sedikit. Jadi dalam penelitian ini, teknik penglihatan mesin telah diterapkan untuk memeriksa permukaan kabel guna mendeteksi kerusakan. Karena kabel berbentuk silinder, dua kamera yang saling berhadapan telah dipasang di atas robot pemanjat tali untuk menutupi seluruh keliling kabel jembatan.

Kedua kamera telah dipicu secara bersamaan untuk memperoleh gambar permukaan kabel. Untuk meningkatkan kinerja pergerakan robot di atas kabel, muncul kebutuhan untuk mengurangi beban yang dibawa oleh robot, sehingga gambar yang diambil telah dikirim ke cloud melalui papan IoT Node MCU untuk diproses lebih lanjut. Penyimpanan cloud juga bertindak sebagai basis data untuk inspeksi dan interpretasi untuk berbagai proses pengambilan keputusan di masa mendatang. Tantangan utama yang dihadapi dalam sistem inspeksi visi mesin adalah variasi pada gambar dengan kondisi pencahayaan yang berbeda, penyok pada bagian penutup, karat dan debu di atas kabel, dll.

Lebih jauh, muncul kebutuhan akan sejumlah besar gambar templat untuk perbandingan dan waktu komputasi yang juga lebih tinggi, yang mungkin bukan metode

terbaik untuk lingkungan yang bergerak secara dinamis. Jadi dalam penelitian ini, jaringan saraf tiruan telah digunakan untuk membuat penilaian, yang dapat mengakomodasi penyimpangan dalam gambar seperti pembentukan skala, rotasi, refleksi, kecerahan dan variasi yang lebih kecil pada gambar karena penyok dan debu.

Sistem yang dikembangkan telah divalidasi dan ditemukan bahwa modul cerdas yang dikembangkan dapat mengidentifikasi bagian kabel jembatan yang rusak secara efektif dan cocok untuk inspeksi komersial waktu nyata. Lebih jauh, kompleksitas komputasi telah dikurangi secara signifikan dengan menggunakan metode SPIHT untuk kompresi gambar tanpa kehilangan data. Lebih jauh, robot yang dikembangkan telah digunakan untuk memeriksa sistem pipa gas medis (MGPS) untuk mengetahui adanya kerusakan dan kebocoran gas.

Berbagai gas yang umum digunakan untuk MGPS adalah Oksigen (O_2), Nitrous oksida (N_2O), Udara medis 400 KPa atau 4 bar / 700 KPa atau 7 bar, Karbon dioksida (CO_2), Nitrogen (N_2), dan Vakum medis. Gas-gas ini digunakan untuk mempertahankan hidup pasien melalui ventilator, mesin anestesi, aplikasi pernapasan, peralatan bedah, tujuan insuficiasi, sistem pembersihan, dll. Pipa-pipa ini diletakkan di luar dan dapat terpapar ke banyak lingkungan berbahaya. Jadi pemeriksaan berkala sangat penting, karena digunakan untuk aplikasi perawatan kesehatan yang menyelamatkan nyawa.

9.2 SISTEM PEMERIKSAAN UNTUK MENDETEKSI CACAT

Secara umum, kabel jembatan terpapar korosi tinggi, suhu tinggi, kelembapan, beban berat, angin, debu, dan pengendapan partikel terkikis, tekanan tinggi, penuaan, curah hujan, dll. Oleh karena itu, kabel dirancang untuk menahan tantangan ini melalui penggunaan kawat baja galvanis, pipa yang dikemas di atas kabel, dll. Cacat apa pun di atas pipa penutup atau kabel akan membuka jalan bagi penyebaran cacat itu ke seluruh kabel. Oleh karena itu, muncul kebutuhan untuk mendeteksi cacat terlebih dahulu.

Di antara berbagai metode untuk mendeteksi cacat, dalam penelitian ini sistem penglihatan mesin telah diadopsi dan berhasil diimplementasikan dengan kinerja yang memuaskan dengan robot pemanjat. Sistem pemeriksaan untuk deteksi cacat pada kabel jembatan telah dikategorikan ke dalam lima submodul berbeda seperti (1) Modul visi mesin, (2) Modul IoT, (3) Modul pemrosesan gambar, (4) Modul ANN, dan (5) Modul robot dalam penelitian ini.

Pada modul visi mesin, dua kamera CCD telah digunakan untuk memperoleh gambar permukaan kabel jembatan. Kedua kamera ini ditempatkan saling berhadapan dan dapat mencakup seluruh keliling kabel. Kamera-kamera ini dipicu secara bersamaan untuk setiap 10 detik atau gerakan robot sejauh 50 cm. Waktu pemicuan bergantung pada kecepatan dan gerakan robot yang dikembangkan. Untuk setiap pemicuan, kamera menangkap gambar kabel sepanjang 50 cm dan itu akan bergantung pada sudut dan panjang fokus kamera.

LED digunakan sebagai sumber cahaya untuk kamera dan LED yang sama telah digunakan untuk memperhatikan lokasi robot oleh pengguna saat berjalan di atas kabel. Tahap berikutnya yang diikuti oleh akuisisi gambar adalah pemrosesan gambar, untuk mengurangi bobot sistem. Dalam penelitian ini, pemrosesan gambar telah dilakukan di sektor jarak jauh,

dengan menyimpan gambar dalam basis data cloud. Tahap kedua adalah modul IoT di mana *Internet of Things* (IoT) memfasilitasi penyimpanan dan pengambilan gambar dari basis data cloud. Papan MCU node dengan modul Wi-Fi ESP 8266 telah digunakan untuk mentransfer gambar ke cloud dan ditunjukkan pada Gambar 9.2a. Ini dikonfigurasi dengan CPU berdaya rendah 32-bit dan standar Wi-Fi. Sinyal gambar dari kamera dikirim ke papan dan gambar-gambar tersebut dapat diberikan ke basis data cloud. Gambar-gambar disimpan dalam basis data dengan tanggal dan waktu. Papan sirkuit IoT ditenagai dengan baterai 12v di robot.

Tajuk
Tajuk Informasi
Opsional Palet
Data Gambar

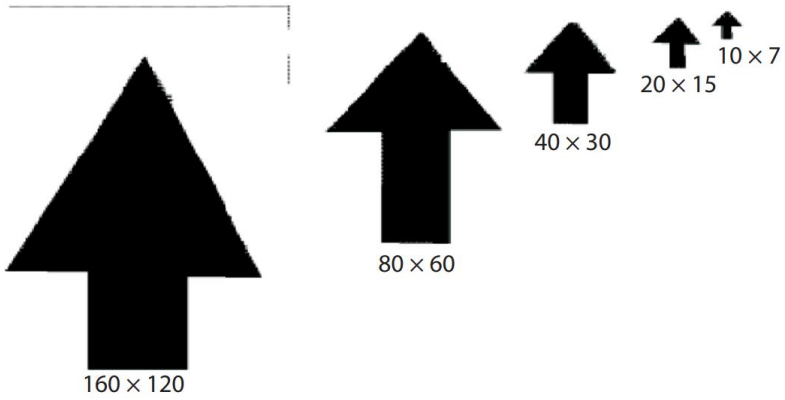
Informasi Tajuk

- Pengidentifikasi Tajuk = 19778
- Ukuran Berkas = 57174
- Dibalik 1 = 0
- Dibalik 2 = 0
- Posisi Awal Data Gambar = 54
- Ukuran Tajuk Informasi = 40
- Lebar Gambar = 159
- Tinggi Gambar = 119
- Jumlah Bidang Warna = 1
- Bit per Piksel = 24
- Rincian Kompresi = 0
- Ukuran Gambar = 57120
- Piksel per Meter pada Sumbu X = 0
- Piksel per Meter pada Sumbu Y = 0
- Jumlah Warna yang Digunakan *(tidak berlaku untuk gambar 24-bit)* = 0
- Jumlah Warna Penting = 0

INFORMASI GAMBAR

202, 198, 198, 200, 195, 196, 198, 194, 192, 205, 201, 199, 214, 207, 206, 210, 204, 203, 207, 201, 1, 203, 220, 212, 20, 60, 59, 48, 63, 60, 49, 63, 60, 49, 63, 62, 51, 63, 62, 51, 64, 63, 52, 64, 63, 52, 65, 64, 53, 65, 64, 53, 6, 63, 54, 65, 63, 54, 0, 200, 199, 190, 200, 199,

190, 200, 199, 190, 200, 199, 190, 200, 199, 190, 200, 199, 190, 200, 199, 190, 200, 199, 190, 200, 199, 19999, 190, 200, 199, 2, 205, 217, 217, 208, 203, 204, 194, 187, 190, 180, 129, 132, 123, 78, 71, 72, 71, 64, 65, 74, 54, 66, 6, 61, 62, 67, 62, 63



Gambar 9.2 Papan Wifi, 2b Format Bitmap dari gambar contoh, 2c metode SPIHT.

Modul ketiga adalah modul pemrosesan gambar. Pada tahap ini, gambar yang diperoleh harus disaring untuk noise dan harus mengurangi ukurannya seminimal mungkin tanpa kehilangan data, sehingga kecepatan pemrosesan ANN meningkat dan waktu komputasi berkurang. Secara khusus, sistem yang dikembangkan membaca gambar input dari cloud dan memprosesnya untuk kembali ke ANN sebagai input. Algoritme ditranskripsi dalam bahasa pemrograman Visual Basic. Modul membaca gambar dari basis data cloud dan mengubah data bitmap warna menjadi data bitmap monokrom dengan mempertimbangkan 0 dan 1 untuk nilai piksel kurang dari 150 dan lebih besar dari 150.

Format de facto sampel yang dibaca oleh perangkat lunak ditunjukkan pada Gambar 9.2b. Dari bit ke-54 bagian data gambar, setiap piksel gambar telah direpresentasikan oleh tiga nilai intensitas warna seperti Biru, Merah dan Hijau, dan rata-rata nilai memberikan data monokrom piksel. Untuk membersihkan noise, setiap piksel telah diekstraksi dan dibandingkan dengan delapan piksel di sekitarnya, kemudian ketidaknormalan telah dihilangkan, seperti yang ditunjukkan pada Gambar 9.3. Sekali lagi, sejumlah besar data raster dalam gambar telah dikurangi menjadi ukuran yang lebih kecil menggunakan metode SPIHT.

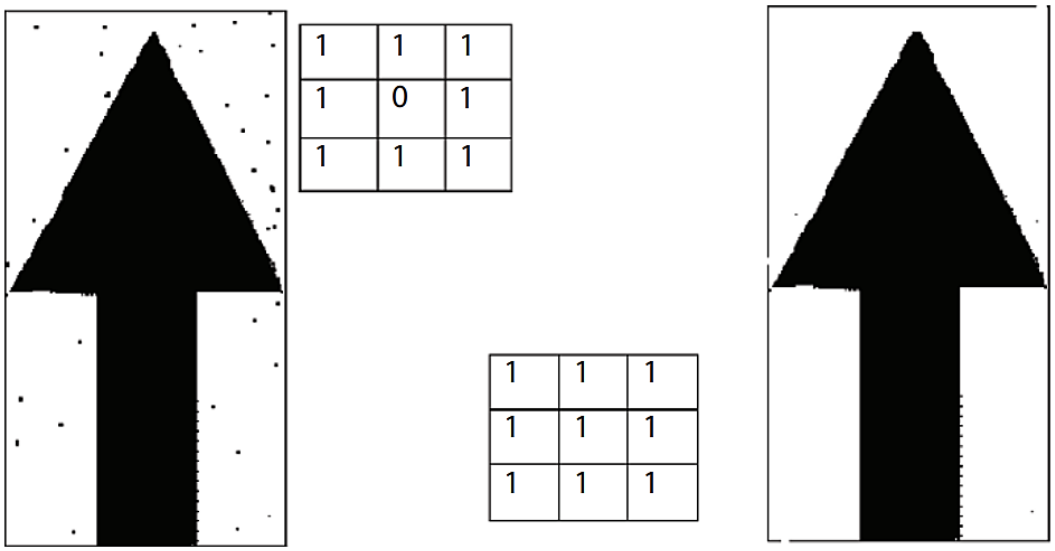
Dalam metode ini, gambar yang difilter harus dibagi menjadi empat kuadran yang sama. Sekali lagi setiap kuadran tunggal telah dibagi menjadi empat kuadran yang sama dan seterusnya. Dalam matriks akhir, kode berulang maksimum akan diantisipasi sebagai nilai untuk matriks tersebut dan pengurangan ukuran sampel ditunjukkan pada Gambar 9.2c. Dalam pendekatan ini, ukurannya telah dikurangi menjadi 70 bit dari 786432 bit. Data 70-bit ini telah diberikan sebagai input ke ANN.

Dari Gambar 9.3, jelas bahwa tingkat noise dalam gambar yang diambil telah dibersihkan. Metodologi model yang dikembangkan ditunjukkan pada Gambar 9.4. Tahap kelima adalah model jaringan saraf tiruan, untuk eksperimen, ANN telah dirumuskan dengan 70,125, 80 dan 21 neuron di lapisan input, tersembunyi dan output, masing-masing. Dalam penelitian ini, nilai kesalahan kuadrat rata-rata yang diharapkan telah ditetapkan menjadi 0,01,

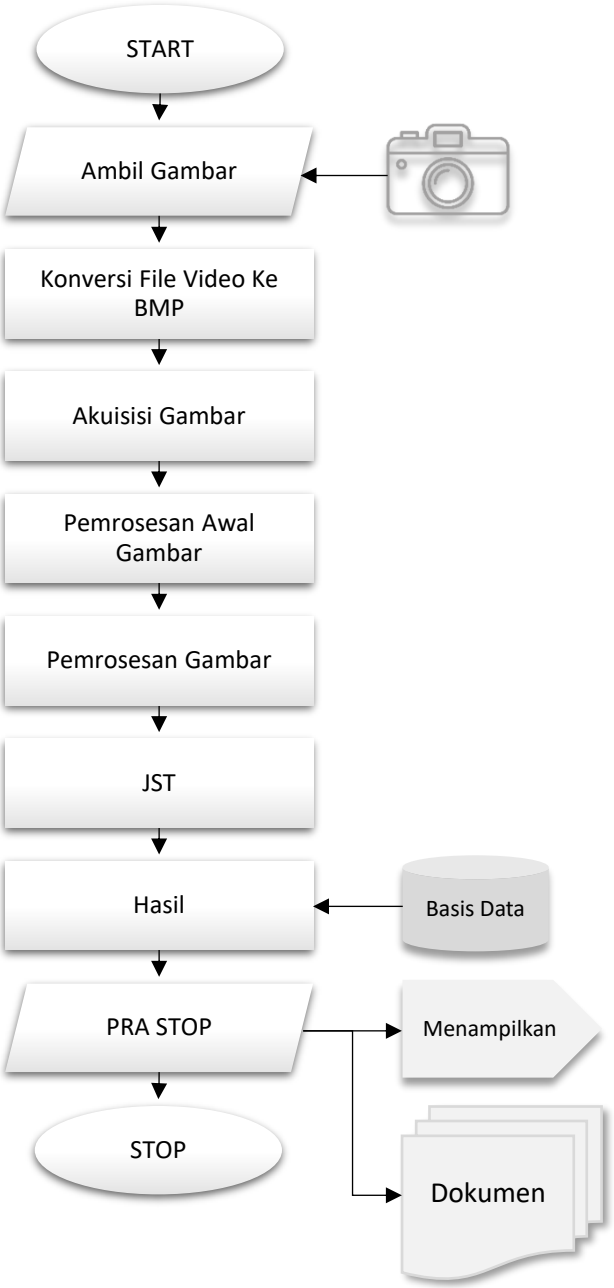
total kesalahan output, yaitu, $\mathrm{T\,error} = \sum \mathrm{Dop} - \mathrm{Oop}$, ditetapkan menjadi 0,01, laju pembelajaran ditetapkan sebagai 0,3. Jaringan yang dikembangkan diberikan pada Gambar 9.5.

Tahap kelima adalah tahap pengembangan robot pemanjat kabel. Ini memiliki modul kereta untuk membawa kamera, baterai, Wi-Fi dan papan kontrol. Gambar 9.5 menunjukkan gambar kereta dan rangka robot. Seluruh kereta telah diperbaiki dengan rangka, yang memiliki roda beralur untuk gerakan. Roda beralur V telah dilengkapi dengan motor penggerak individual untuk bergerak di seluruh kabel ke segala arah bersama dengan pegas kompresi untuk menahan roda dengan kabel, dan roda dapat menggelinding di atas kabel tanpa banyak selip. Ukuran alur roda harus diatur berdasarkan diameter kabel, untuk kinerja yang unggul. Ukuran roda yang dipertimbangkan untuk eksperimen memiliki diameter 150 mm. Robot yang dikembangkan telah dipelajari dengan dinamikanya untuk keseimbangan, sehingga kinerja pendakian ditingkatkan.

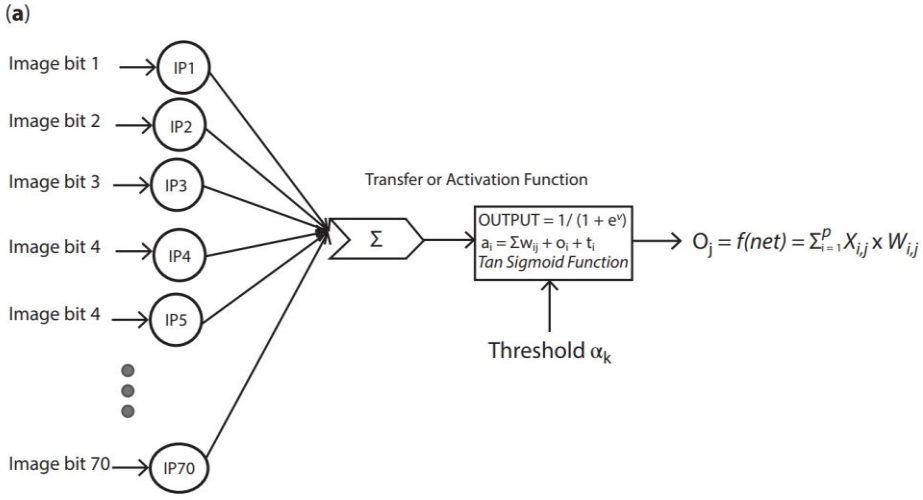
Pada tahap pertama, gambar diambil menggunakan kamera digital sebagai gambar bitmap. Pada tahap pemrosesan, data bitmap warna diubah menjadi data bitmap monokrom. Derau dalam data difilter menggunakan algoritma filter. Pada tahap pemrosesan, ukuran data gambar diubah menjadi ukuran yang dapat diterima oleh algoritma BPN yang telah dilatih menggunakan SPIHT. Data yang telah diproses ini dikirim ke ANN untuk pemeriksaan gambar. Kemudian keluaran ANN diinterpretasikan untuk pemeriksaan komponen. Tahap-tahap ini dijelaskan secara rinci di bagian berikut.

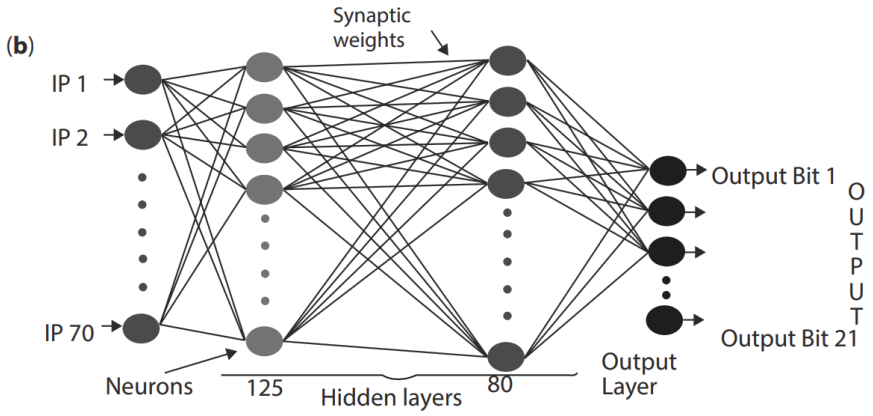


Gambar 9.3 Pengurangan kebisingan.



Gambar 9.4 Metodologi model yang dikembangkan.





Gambar 9.5 Jaringan ANN.

Pada tahap pertama, gambar diambil menggunakan kamera digital sebagai gambar bitmap. Pada tahap pemrosesan, data bitmap warna diubah menjadi data bitmap monokrom. Derau dalam data difilter menggunakan algoritma filter. Pada tahap pemrosesan, ukuran data gambar diubah menjadi ukuran yang dapat diterima oleh algoritma BPN yang telah dilatih menggunakan SPIHT. Data yang telah diproses ini dikirim ke ANN untuk pemeriksaan gambar. Kemudian keluaran ANN diinterpretasikan untuk pemeriksaan komponen. Tahap-tahap ini dijelaskan secara rinci di bagian berikut.

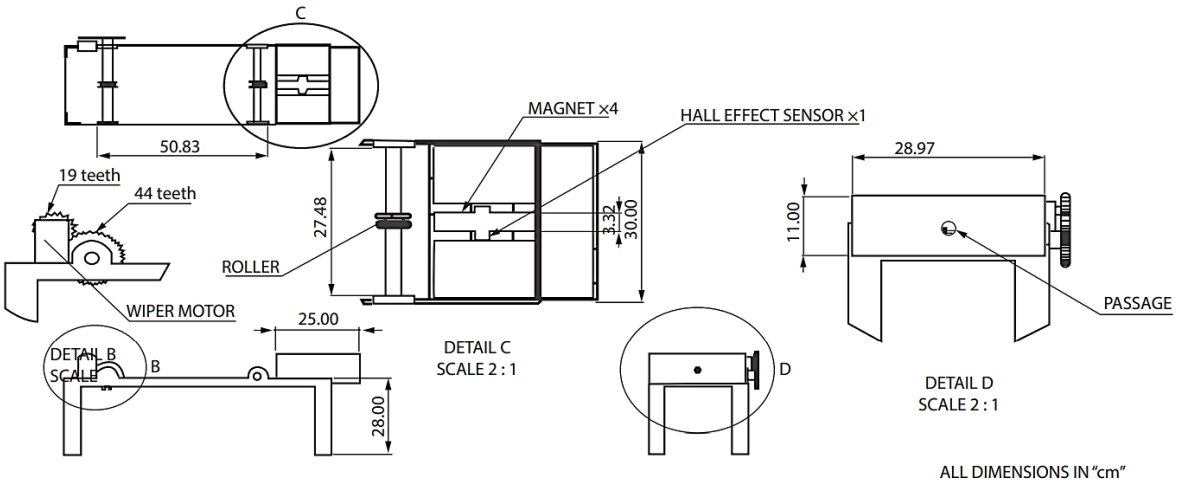
9.3 METODOLOGI PENGENALAN CACAT

Komponen penting untuk sistem penglihatan mesin adalah sumber cahaya; LED telah digunakan sebagai sumber cahaya karena kecerahannya dapat dikontrol dengan mudah. LED juga digunakan sebagai indikator untuk menemukan robot di atas kabel ketinggian tinggi. Kamera CCD adalah bagian inti untuk sistem deteksi kerusakan. Saat robot pemanjat merangkak di atas kabel jembatan, untuk setiap gerakan 50 cm, pengontrol Arduino mengirimkan sinyal pemicu ke kamera. Segera setelah menerima perintah, kamera menangkap gambar permukaan kabel dan gambar permukaan yang diperoleh dikirim ke basis data cloud melalui modul IoT. Kamera ditempatkan tepat berlawanan satu sama lain dan pada panjang fokus yang sama. Bingkai tunggal gambar kamera beresolusi 1024 x 768 piksel dan disimpan dalam format bitmap untuk pemrosesan yang mudah. Gambar diunggah ke cloud dan cloud memperbarui data pada waktu yang telah ditentukan sebelumnya.

Kemudian modul pemrosesan gambar yang dikembangkan memproses gambar yang disimpan di cloud di lokasi terpencil. Berbagai tahap pemrosesan gambar adalah pemotongan tepi, penyaringan data yang berisik, konversi ke gambar monokrom untuk mengurangi 1/3 dari ukurannya, konversi ke gambar biner untuk memudahkan pemrosesan oleh ANN dan pengurangan ukuran tanpa kehilangan data atau kehilangan bentuk menggunakan partisi set dalam metode pohon Hierarki hingga 28 kali. Ukuran akhir gambar berisi 70 digit biner. 70 digit biner ini telah diberikan sebagai masukan ke modul ANN terlatih yang dikembangkan. Jumlah neuron dalam lapisan masukan adalah 70, diikuti oleh dua lapisan tersembunyi yang terdiri dari 125 dan 80 neuron dan lapisan keluaran dengan 21 neuron.

Contoh gambar dengan urutan yang dikodekan ANN dan keputusan untuk urutan ANN ditunjukkan pada Gambar 9.7. Sebelum implementasi, sampel 150 gambar kabel dan pipa penutup telah dikumpulkan, 125 gambar telah digunakan untuk melatih jaringan saraf tiruan dan 25 gambar telah digunakan untuk menguji jaringan. Untuk melatih algoritma backpropagation jaringan dengan fungsi sigmoid tan telah digunakan. Algoritma backpropagation telah memodifikasi nilai ambang neuron dan nilai bobot tautan dari lapisan belakang ke lapisan depan berdasarkan nilai kesalahan yang diperoleh. Persamaan dan rumus yang digunakan untuk pembaruan diberikan pada Gambar 9.4.

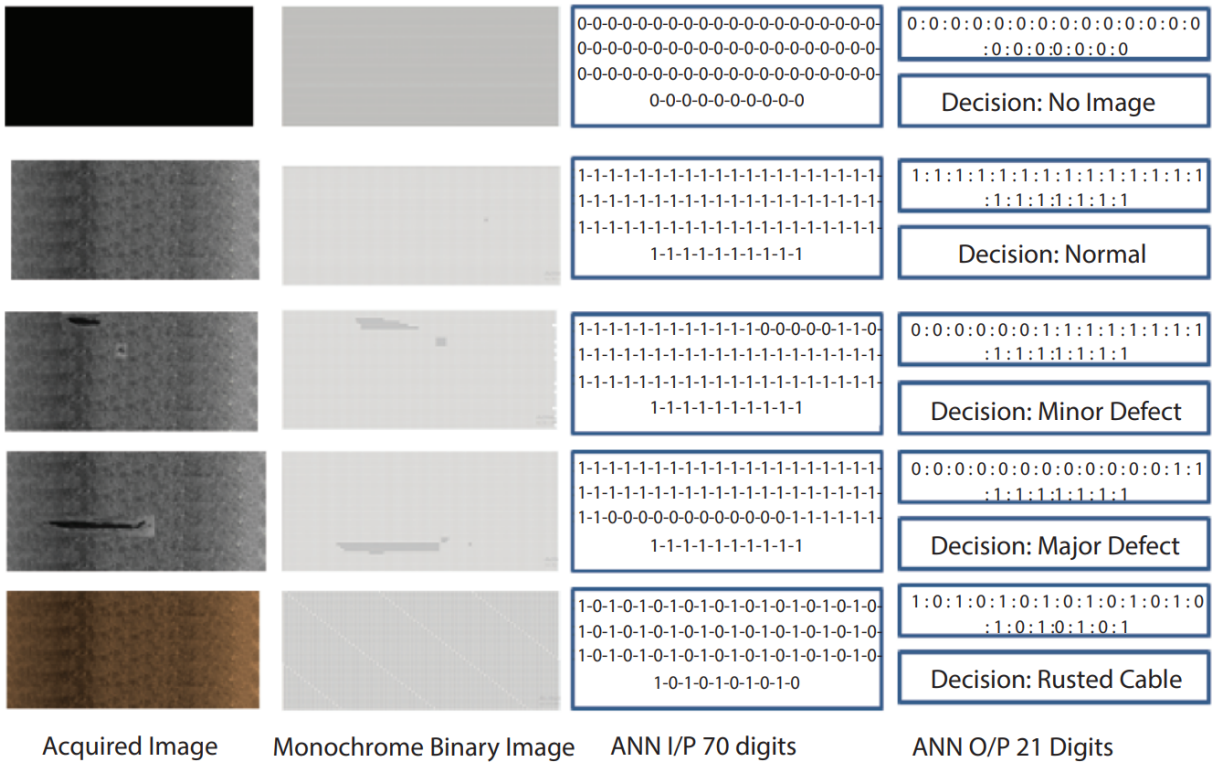
Awalnya nilai ambang dan bobot diasumsikan sebagai nilai acak. Bobot akhir yang dilatih dan nilai ambang disimpan dalam file notepad dan disimpan lagi di cloud. Modul ANN yang dilatih ini memproses data yang diberikan dan nilai yang diperoleh pada lapisan keluaran harus dikodekan ke format yang dapat dipahami manusia. Kombinasi contoh keluaran dan nilai keputusan telah ditunjukkan pada Gambar 9.6. Untuk berbagai kombinasi 21 digit biner, keputusan yang didekode adalah tidak ada gambar, gambar tidak diketahui, tidak ada cacat, cacat kecil atau debu, kabel rusak, dan kabel berkarat. Berbagai jenis contoh gambar dan keputusan ditunjukkan pada Gambar 9.7.



Gambar 9.6 Kereta dan rangka robot

.

Percobaan telah dilakukan di laboratorium dengan berbagai jenis kabel dan pipa tertutup pada berbagai sudut dan orientasi. Hasil yang diberikan oleh robot cukup memuaskan. Jadi modul yang dikembangkan dapat digunakan untuk pemeriksaan kabel jembatan. Tantangan utama yang dihadapi dalam implementasi adalah kecepatan pendakian robot yang berubah-ubah pada berbagai orientasi, pengunggahan gambar secara berkala ke cloud, dan waktu komputasi lebih lambat daripada waktu pengambilan, sehingga prediksi langsung tidak dapat dilakukan. Model yang dikembangkan dapat digunakan untuk aplikasi lain juga. Jadi model yang dikembangkan digunakan untuk aplikasi perawatan kesehatan.



Gambar 9.7 Contoh gambar dan keputusan ANN.

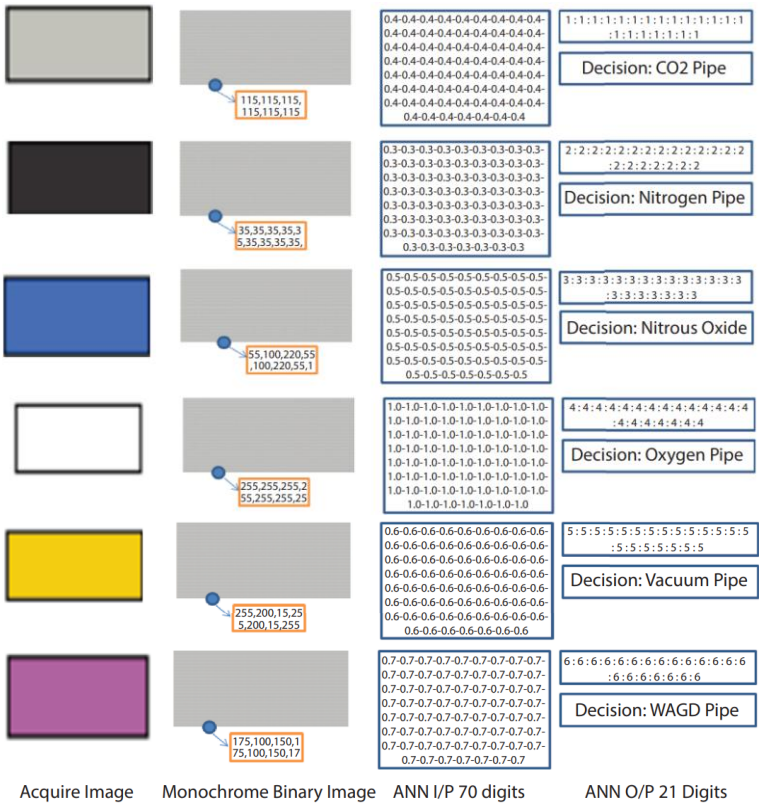
9.4 INSPEKSI MGPS PERAWATAN KESEHATAN

Di sebagian besar rumah sakit, MGPS umumnya digunakan untuk berbagai aktivitas penyelamatan nyawa. Pipa-pipa ini juga bertekanan tinggi dan terpapar lingkungan luar. Contoh sistem perpipaan ditunjukkan pada Gambar 9.8. Robot yang dikembangkan dapat digunakan untuk mendeteksi kerusakan pada sistem perpipaan secara berkala. Biasanya pipa yang digunakan untuk MGPS adalah pipa tembaga dan memiliki warna yang berbeda untuk memudahkan identifikasi. Warna dan gas di dalam pipa diberikan pada Gambar 9.8. Robot telah diprogram untuk kode warna yang diberikan pada Gambar 9.9.

Gambar 9.8 Contoh MGPS.

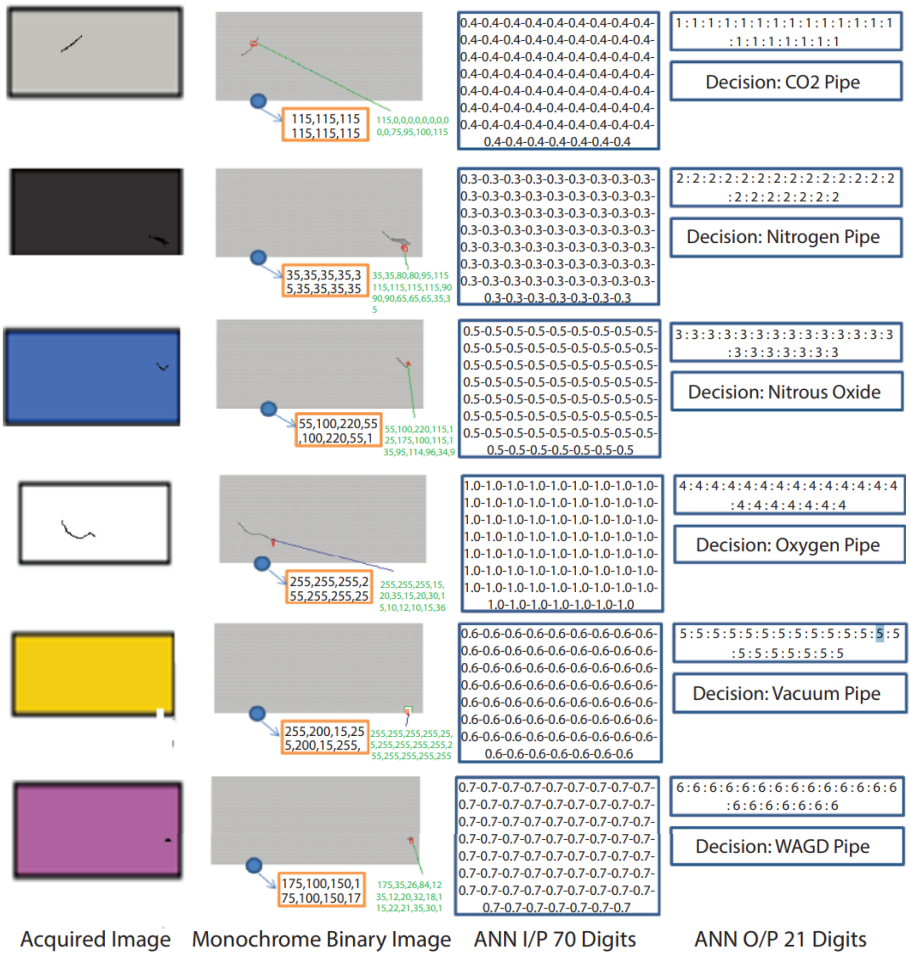
GAS	WARNA ISO
Karbon Dioksida	Abu-abu
He-O_2	Coklat Putih
Instrumen Udara	
Udara Medis	Hitam Putih
Nitrogen	Hitam
Dinitrogen Oksida	Biru
O_2-He	Putih Coklat
Oksigen	Putih
Vakum (Hisab)	Kuning
WAGD (Evakuasi)	Ungu

Gambar 9.9 MGPS; kode warna.

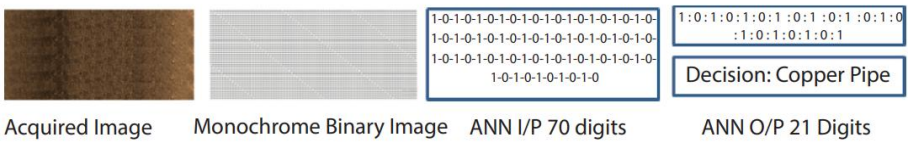


Gambar 9.10 Gambar dan penyandian ANN untuk pipa berwarna.

Robot yang dikembangkan telah diprogram untuk beroperasi dalam tiga modul berbeda. Pada modul pertama, berdasarkan kode warna pada pipa gas, robot mendeteksi jenis gas di dalamnya dan memberi tahu pengguna. Pada modul kedua, pada pipa gas terkait, kerusakan atau cacat apa pun akan diidentifikasi. Di beberapa rumah sakit, karena alasan ekonomi, kode warna hanya diberikan di kamar pasien, di luar dengan tabung tembaga. Jadi pada modul ketiga, robot yang dikembangkan dapat mendeteksi cacat pada tabung tembaga. Berbagai gambar yang diperoleh dan mekanisme deteksi diberikan pada Gambar 9.10 dan Gambar 9.12. Gambar 9.10 menjelaskan berbagai kode warna dan input dan output ANN untuk pipa gas yang berbeda. Gambar 9.11 menunjukkan cacat pada pipa berkode warna. Gambar 9.12 menunjukkan pipa tembaga dan representasi digit yang dikodekan.

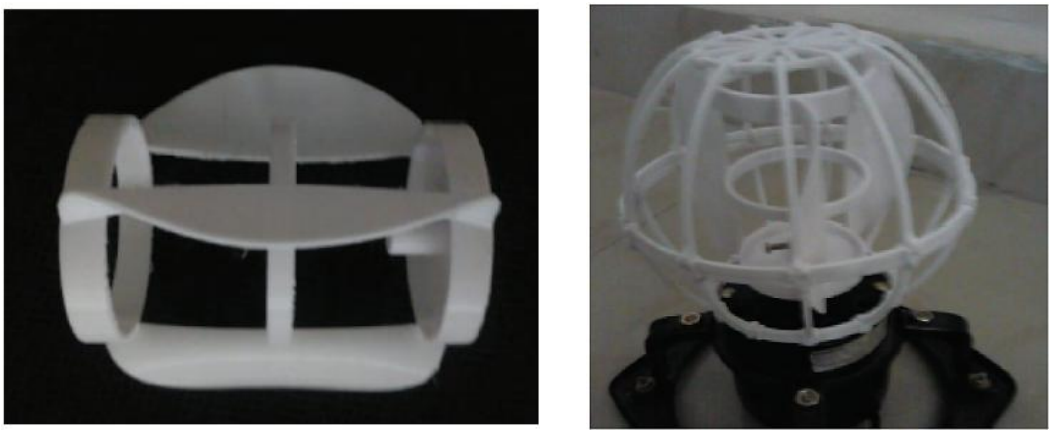


Gambar 9.11 Gambar cacat dan penyandian ANN untuk pipa berwarna.



Gambar 9.12 Gambar pipa tembaga dan penyandian ANN.

Meskipun robot mendeteksi cacat pada saluran pipa, kebocoran juga terdeteksi dengan adanya kipas dan mekanisme generator. Setiap kebocoran gas akan memutar bilah kipas dan dengan demikian generator menghasilkan emf; setelah emf dibaca oleh mikrokontroler, sinyal peringatan diberikan kepada pengguna.



Gambar 9.13 Kipas pengurang kebocoran.

Gambar bilah kipas yang dikembangkan untuk mendeteksi kebocoran gas ditunjukkan pada Gambar 9.13. Bilah dibuat menggunakan printer 3D dan sifat khususnya ringan dan dapat berputar saat gas mengenai area mana pun pada permukaan bilah. Bilah kipas berbeda dari bilah komersial.

Dalam Bab ini, robot inspeksi cerdas berbasis penglihatan untuk inspeksi kabel jembatan telah dikembangkan dengan teknologi IoT untuk menyimpan data di cloud untuk diproses. Robot telah digunakan untuk mendeteksi kerusakan permukaan pada kabel jembatan. Cacat permukaan pada kabel jembatan dapat secara otomatis dideteksi dan diberitahukan ke sistem jarak jauh untuk tindakan perawatan yang diperlukan. Sementara itu, SPIHT yang disederhanakan dapat diimplementasikan secara efisien untuk pengurangan ukuran tanpa banyak kehilangan data. Hasil investigasi menyimpulkan bahwa akuisisi gambar, pemrosesan gambar, dan identifikasi cacat telah dilakukan secara efisien untuk deteksi kerusakan permukaan kabel/tali jembatan oleh model yang dikembangkan. Pekerjaan selanjutnya dapat dilakukan pada pemrosesan waktu nyata, peningkatan gerakan robot di atas kabel, dan pemrosesan paralel untuk mengurangi waktu komputasi. Robot yang dikembangkan juga telah digunakan untuk menguji sistem pipa gas medis untuk mengetahui adanya cacat dan kebocoran.

BAB 10
PENDEKATAN FUZZY UNTUK MENDESAIN

Keamanan siber bukan hanya satu kesulitan; biasanya ini adalah masalah yang melibatkan banyak aspek berbeda. Objek/sistem berbasis Aturan Fuzzy untuk keamanan siber dapat berupa sistem yang terdiri dari kumpulan aturan dan mekanisme untuk mengakses dan menjalankan prinsip-prinsip tersebut. Kumpulan aturan biasanya dibangun dengan serangkaian kelompok/set aturan terkait. Dampak aktivitas kriminal siber bergantung pada karakter kejahatan dan sifat korban. Arab Saudi menghadapi banyak ancaman siber termasuk *Denial of Service* (DoS), malware, pencemaran nama baik situs web, dan serangan email spam dan phishing.

Meskipun temuan baru-baru ini menyoroti buruknya sistem keamanan informasi Arab Saudi, dalam premis saat ini disarankan agar penilaian risiko keamanan siber khusus dapat dilakukan. Dengan menggunakan teori logika Fuzzy, kami mengusulkan Model Inferensi Fuzzy (FIS) untuk menyediakan mitigasi risiko dan memeriksa untuk mengurai masalah tersebut kepada entitas yang diusulkan. Tujuan dari lokasi gangguan adalah untuk mengamati latihan yang terorganisir dan akibatnya, mengenali serangan balas dendam dan memutuskan desain keamanan pengaturan PC yang tepat.

10.1 PENDAHULUAN

Saat ini asosiasi, yang melihat dengan cakupan yang layak dari kemungkinan Ancaman terhadap keamanan data (IS) mereka, semakin ingin tahu tentang tingkat signifikannya. Salah satu pendekatan yang paling mudah untuk mengukur, mencapai, dan memelihara keamanan informasi adalah tinjauan Keamanan Informasi. Tinjauan keamanan (diperiksa secara menyeluruh) mungkin merupakan pekerjaan yang membingungkan, banyak tahap, dan prosedur serius yang melibatkan profesional (spesialis) yang sangat berkualifikasi dalam IS, yang membuatnya sangat mahal.

Kerentanan memengaruhi dinamika; gagasan informasi secara inheren diidentifikasikan dengan gagasan kerentanan. Bagian utama pertama dari asosiasi ini adalah bahwa kerentanan yang terlibat dengan keadaan berpikir kritis apa pun mungkin merupakan konsekuensi dari beberapa kekurangan data, yang dapat tidak memadai, tidak pasti, terpisah-pisah, tidak sepenuhnya solid, ambigu, berlawanan, atau kurang dalam hal lain. Kerentanan adalah sifat informasi. Karena IoT, di masa depan individu akan memiliki fondasi pemrosesan yang tidak terdeteksi dan ada di mana-mana untuk melakukan berbagai latihan baik di tempat kerja maupun di rumah. Rumah masa kini membutuhkan gadget yang mudah digunakan dan bersinergi. Pada awalnya, kontrol rasional representatif dikenal dengan pendekatan konfigurasi kontrol bebas model tetapi disensor karena tidak adanya investigasi keamanan dan rencana regulator yang efisien.

1. Cakupan dan studi pra-tinjauan: menentukan zona fokus yang paling; menyiapkan tujuan tinjauan.
2. Pengaturan dan persiapan: sebagian besar menghasilkan rencana/agenda kerja tinjauan.

3. Pekerjaan langsung: mengumpulkan bukti dengan bertemu staf dan administrator, menilai arsip, cetakan dan informasi, mengamati formulir dalam kehidupan nyata, dan sebagainya.
4. Investigasi: mengamati, memeriksa, dan menganalisis bukti-bukti yang terkumpul mengenai tujuan.
5. Pengumuman: menjelajahi setiap tahap sebelumnya, menemukan hubungan dalam data yang terkumpul, dan membuat laporan.
6. Kesimpulan. Setiap tahap memiliki jumlah data yang sangat besar, yang harus dicatat, disortir, dan akhirnya dipecah.

Kemajuan pesat inovasi informasi, serangan sistem dan PC telah memicu kekhawatiran luas di seluruh dunia. Tidak hanya terjadi peningkatan jumlah dan jenis serangan, sifat dan kelas yang beraneka ragam juga telah diperluas. Kemungkinan kerusakan akibat serangan semakin nyata. Karena keamanan Internet mungkin merupakan bidang yang bergerak cepat, serangan yang mendapatkan fitur dapat berubah secara mendasar dari satu tahun ke tahun berikutnya.

Sistem Pakar Fuzzy (FES) yang didukung oleh pemikiran terukur dicirikan dan diandalkan untuk mengerjakan transportasi pesaing yang wajar selama kerangka dispersi untuk situasi kapasitor. Tegangan dan catatan penurunan kehilangan daya dari transportasi kerangka penyebaran ditunjukkan oleh kapasitas pendaftaran Fuzzy. Sistem Pakar Fuzzy yang berisi kumpulan standar heuristik kemudian digunakan untuk mengerjakan kesesuaian pengaturan kapasitor setiap transportasi di dalam kerangka pengiriman. Kapasitor kemudian diposisikan pada hub dengan kewajaran terbaik yang absolut dengan memanfaatkan kemajuan berikut:

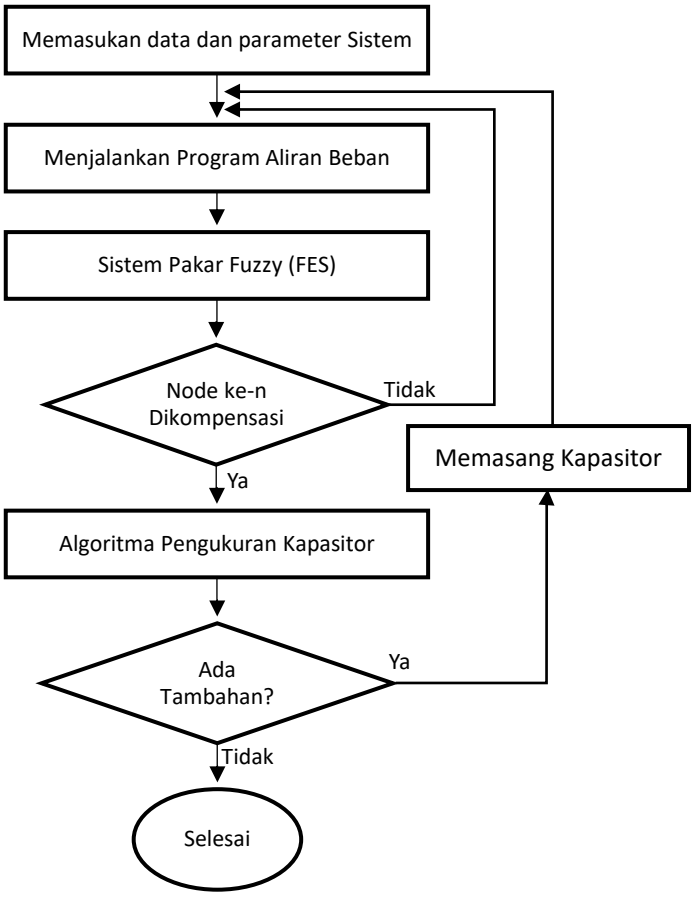
- Program aliran beban menghitung penurunan kehilangan daya dari kerangka kerja mengenai setiap transportasi ketika arus beban reseptif pada transportasi itu diperbaiki. PLR kemudian secara langsung distandarisasi ke dalam wilayah, dengan penurunan kehilangan daya absolut menjadi 1 dan dengan demikian yang terendah menjadi 0.
- Catatan penurunan kerugian daya di dekat tegangan nodal per unit (sebelum pembayaran PLR) adalah dua sumber data yang digunakan oleh Sistem Pakar Fuzzy. FES kemudian memutuskan hub utama yang sesuai untuk pemasangan kapasitor dengan deduksi Fuzzy.
- Pekerjaan penghematan biaya mengenai biaya tahunan kapasitor (dikendalikan oleh ukuran) dicirikan dan dengan demikian penghematan yang diperoleh dari penurunan kerugian daya diperkuat untuk menghitung estimasi ideal bank kapasitor yang akan dipasang di hub itu.
- Prosedur di atas akan diulang, mencari hub lain untuk posisi kapasitor, hingga tidak ada lagi dana cadangan anggaran yang sering ditemukan.
    1. Penataan dan persiapan: biasanya membuat rencana kerja/agenda tinjauan.
    2. Pekerjaan langsung: mengumpulkan bukti dengan berbicara kepada staf dan supervisor, memeriksa laporan, cetakan dan informasi, melihat formulir di dunia nyata, dan sebagainya.
    3. Pemeriksaan: melihat, memeriksa dan menganalisis bukti yang dikumpulkan mengenai tujuan.

4. Perincian: memeriksa setiap tahap sebelumnya, menemukan hubungan dalam data yang dikumpulkan dan membuat laporan.

Alasan fuzzy mungkin merupakan alasan yang sangat dihargai atau alasan probabilistik; ia mengelola pemikiran yang diperkirakan daripada yang tetap dan pasti. Menariknya dengan alasan adat, mereka akan memiliki nilai yang berbeda, di mana himpunan ganda memiliki alasan dua nilai, valid atau palsu, faktor alasan simbolik mungkin memiliki nilai realitas yang berkisar dalam derajat di suatu tempat dalam kisaran 0 dan 1.

Alasan simbolik telah direntangkan untuk menangani gagasan kebenaran fraksional, di mana nilai kebenaran dapat berkisar antara yang sepenuhnya jelas dan yang sepenuhnya salah. Selain itu, ketika faktor fonetik digunakan, derajat ini juga dapat dikelola oleh kapasitas eksplisit. Strategi alasan representatif telah digunakan di dalam bidang keamanan komputer sejak tahun 1990-an. Alasan representatif juga mencakup potensi yang ditunjukkan di dalam bidang penemuan gangguan dibandingkan dengan kerangka kerja yang menggunakan pencocokan tanda yang parah atau pengenalan penyimpangan contoh teladan.

Gagasan ketidakjelasan membantu menghaluskan pemisahan tiba-tiba perilaku normal dari perilaku tidak teratur. Alasan representatif mencakup kapasitas untuk berbicara dengan berbagai jenis pemikiran yang longgar di wilayah-wilayah di mana keputusan tegas harus dibuat dalam situasi yang tidak meyakinkan seperti pengenalan gangguan. Sistem pakar fuzzy harus dijelaskan dalam Gambar 10.1 (R. Chandia *et al.*, 2007).



Gambar 10.1 Sistem pakar Fuzzy.

10.2 HIMPUNAN FUZZY

1. Himpunan fuzzy mungkin merupakan himpunan yang memiliki derajat partisipasi di suatu kisaran 1 dan 0. Himpunan fuzzy dilambangkan dengan karakter tilde (~). Misalnya, jumlah kendaraan yang mengikuti lampu lalu lintas pada jarak tertentu dari semua kendaraan yang hadir akan memiliki nilai partisipasi antara.
2. Partisipasi parsial terjadi ketika individu dari 1 Himpunan Fuzzy juga dapat menjadi bagian dari Himpunan Fuzzy lainnya di dalam alam semesta yang sama.
3. Tingkat partisipasi atau kebenaran tidak sama dengan probabilitas. Kebenaran fuzzy mengacu pada pendaftaran dalam himpunan yang dicirikan secara ambigu.

Pendahuluan Himpunan Fuzzy Pada bagian ini, konsep Himpunan Fuzzy dan akibatnya prosedur pada Himpunan Fuzzy dibahas. Konsep tersebut merupakan spekulasi dari himpunan baru. Himpunan lama juga disebut himpunan 'baru' sehingga kita dapat mengenalinya dari Himpunan Fuzzy. Sejujurnya, himpunan Crisp sering dianggap sebagai contoh unik dari himpunan Fuzzy. Biarkan An sebagai himpunan baru yang dikarakterisasi di Semesta X. Pada saat itu untuk setiap komponen x di X, baik x mungkin merupakan individu dari An atau tidak. Dalam ilmu murni Fuzzy, properti ini diringkas. Oleh karena itu, selama himpunan Fuzzy, agak berlebihan jika x mungkin merupakan Anggota penuh dari himpunan atau bukan bagian. Sering kali merupakan individu setengah jalan dari himpunan yang disebutkan dalam Gambar 10.2.

Spekulasi dilanjutkan sebagai berikut: Untuk setiap himpunan baru An, dapat dibayangkan untuk mengkarakterisasi Kapasitas Karakteristik atau pekerjaan pendaftaran μP = {0, 1}. yaitu, pekerjaan merek dagang mengambil kedua kualitas 0 atau 1 di dalam himpunan tradisional. Untuk himpunan Fuzzy, kapasitas merek dagang dapat mengambil insentif apa pun di suatu tempat dalam kisaran nol dan satu.

Definisi:

Pekerjaan pendaftaran μP(x1) dari himpunan Fuzzy A mungkin merupakan kapasitas μP: X1→[‘benar’, ’salah’] Jadi, setiap komponen dalam x1 di X memiliki derajat partisipasi: μA(x1) ∈ [‘benar’, ’salah’] An sepenuhnya dikendalikan oleh susunan tupel: P = {(x1, μP(y)) x∈ X}.

Contoh:

Asumsikan seseorang perlu mengklarifikasi objek kendaraan yang memiliki properti mahal dengan mempertimbangkan BMW, Rolls Royce, Mercedes, Ferrari, Fiat, Honda, dan Renault. Beberapa kendaraan seperti Ferrari dan Rolls Royce tidak diragukan lagi mahal dan beberapa, seperti Fiat dan Renault, tidak mahal jika dibandingkan dan tidak termasuk dalam himpunan tersebut. Dengan menggunakan himpunan Fuzzy, susunan Fuzzy kendaraan mahal sering digambarkan sebagai:

himpunan Fuzzy tidak dibedakan dari susunan logika Boolean itu sendiri dengan kapasitas partisipasi tambahan di antara "sah" dan "salah". Seperti namanya, itu adalah metode dasar logika berpikir yang diperkirakan daripada yang pasti. Pentingnya logika simbolik berasal dari kenyataan bahwa sebagian besar metode berpikir manusia dan khususnya pemikiran indrawi diperkirakan di alam.

Gambar 10.2 Set yang tajam vs. Set Fuzzy.

Atribut fundamental dari logika emblematik adalah sebagai berikut:

- Dalam logika emblematik, pemikiran akurat dilihat sebagai contoh terbatas dari pemikiran terukur.
- Dalam logika emblematik, semuanya mungkin melibatkan derajat.
- Setiap logika biasanya dikaburkan.
- Dalam logika emblematik, informasi diuraikan sebagai banyak hal yang fleksibel atau, identik dengan, keharusan Fuzzy pada banyak faktor.
- Inferensi dilihat sebagai prosedur penyebaran batasan fleksibel.
- Ada dua kelas kejahatan yang signifikan dengan PC:
- Penggunaan PC yang tidak sah, yang dapat mencakup pengambilan nama pengguna dan kata sandi, atau dapat mencakup akses ke PC korban melalui web melalui jalur sekunder yang dikerjakan oleh program bug.
- Membuat atau mengirimkan infeksi PC ganas (misalnya, infeksi PC, worm, Trojan horse).

Ketika orang mendengar kata-kata "Kesalahan sistem", mereka sering kali menganggap gambar-gambar tidak senonoh dapat diakses di web. Masalah optimal di web biasanya sebanding dengan fakta yang berguna tentang ketidaksenonohan dalam buku, selain dari berbagai masalah logis dengan cakupan tunggal di situs tersebut. Kesalahan umum yang melibatkan Sistem sama dengan pelanggaran tanpa Sistem. Sistem hanyalah instrumen yang digunakan kasus penipuan untuk melakukan pelanggaran penipuan. Pertimbangkan hal berikut:

- menggunakan Sistem, pemindai, pemrograman desain, dan printer warna yang bagus untuk meniru atau memalsukan adalah kesalahan yang sama seperti menggunakan mesin cetak lama yang bagus dengan tinta.
- Mencuri sistem dengan data eksklusif yang disimpan di pelat keras di dalam sistem adalah kesalahan yang sama seperti mencuri tas.
- Layanan web untuk meminta seks hampir sama dengan jenis penjualan lainnya, pada saat itu bukanlah kesalahan penggantian.
- Sistem sering kali sebaliknya untuk mengirimkan pencurian atau pemalsuan.

Prosedur pada himpunan Fuzzy

Tugas penting yang mungkin dilakukan pada himpunan Fuzzy adalah aktivitas asosiasi, konvergensi, suplemen, item matematika, dan keseluruhan aritmatika. Banyak eksplorasi

mengenai himpunan Fuzzy dan aplikasinya pada hipotesis automata, alasan, kontrol, permainan, geografi, pengakuan desain, sangat penting, etimologi, klasifikasi ilmiah, kerangka kerja, pemilihan, pemulihan data, dan sebagainya, telah dianut dengan tulus dengan memanfaatkan aktivitas ini untuk himpunan Fuzzy.

Selain aktivitas tersebut, tugas baru yang disebut "agregat terbatas" dan terlebih lagi pada aktivitas tersebut, tugas baru yang disebut "keseluruhan terbatas" dan "perbedaan terbatas" disajikan oleh Zadeh (1975) untuk mengeksplorasi pemikiran Fuzzy yang menjelaskan cara menangani masalah pemikiran yang terlalu membingungkan untuk pengaturan yang tepat. Berbagai jenis operator:

a) Keseimbangan
b) Suplemen
c) Titik perpotongan
d) Asosiasi
e) Item matematika
f) Augmentasi himpunan Fuzzy dengan Angka Crisp
g) Intensitas himpunan Fuzzy (viii) Agregat matematika

1. Himpunan Fuzzy ekuivalen :

Kita harus mempertimbangkan dua himpunan Fuzzy $P(y)$ dan $Q(y)$ yang dianggap ekuivalen, jika $\mu P(y) = \mu Q(y)$ untuk semua $x \in X$. Hal ini dikomunikasikan sebagai berikut:

$$P(y) = Q(y), \text{if } \mu P(y) = \mu Q(y)$$

Catatan:

Dua himpunan Fuzzy $P(y)$ dan $Q(y)$ dianggap tidak konsisten, j*ika* $\mu P(y) \neq \mu Q(y)$ *un*tuk setidaknya $y \in Y$.

Contoh:

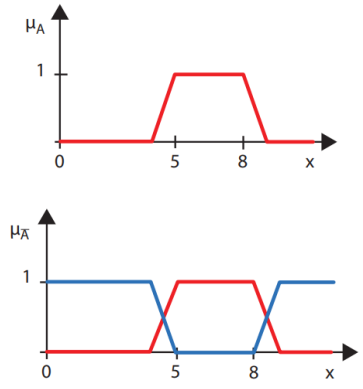
$$P(y) = \{(y1,0.1), (y2,0.2), (y3,0.3), (y4,0.4)\}$$
$$Q(y) = \{(y1,0.1), (y2,0.2), (y3,0.3), (y4,0.6)\}$$

Karena $\mu P(y) \neq \mu Q(y)$ untuk $y \in Y$ yang berbeda, maka $P(y) \neq Q(y)$

2. Suplemen himpunan Fuzzy $P(y)$:

Suplemen adalah sesuatu yang berlawanan dengan himpunan tersebut. Suplemen himpunan Fuzzy dilambangkan dengan $p'(y)$ dan dicirikan seperti pada himpunan inklusif Y seperti yang ditunjukkan pada Gambar 10.3 berikut:

$$P'(y) = 1 - P(y) \text{for all } y \varepsilon Y$$



Gambar 10.3 Contoh operasi komplemen pada himpunan fuzzy.

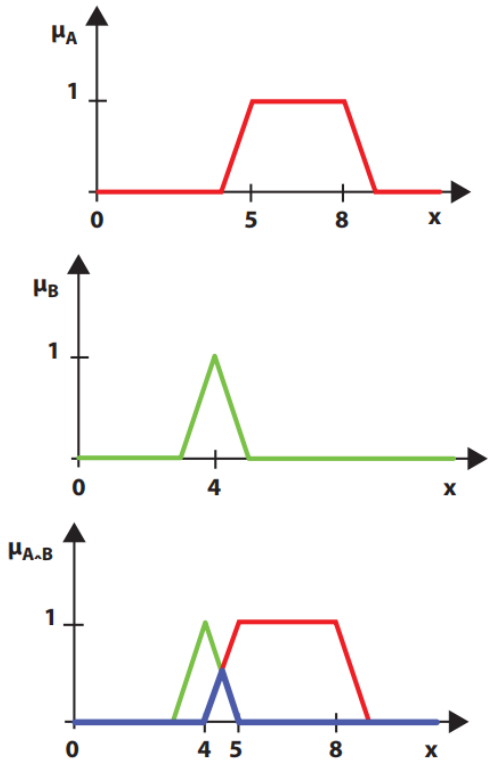
3. Titik perpotongan himpunan fuzzy :

Segmen himpunan fuzzy mencirikan sejauh mana komponen tersebut berada dalam dua himpunan. Dapat memiliki berbagai tingkat partisipasi dalam setiap himpunan. Tingkat partisipasi adalah pendaftaran terendah dalam dua susunan setiap komponen. Misalkan $P(y)$ dan $Q(y)$ adalah dua himpunan fuzzy, konvergensi dari didefinisikan sebagai $(P \cap Q)(y)$ dan dengan demikian nilai kerja partisipasi diberikan sebagai berikut:

$$\mu(p \cap Q)(y) = \min\{\mu Q(y)\}$$

Titik persimpangan secara praktis setara dengan aktivitas AND yang masuk akal.

$$P(y) = \{y1,0.7), (y2,0.3), (y3,0.9), (y4,0.1)\}$$
$$Q(y) = \{(y1,0.2), (y2,0.5), (y3,0.7), (y4,0.4)\}$$
$$\mu(P \cap Q)(y1) = \min\{\mu P(y1), \mu Q(y1)\} = \min\{0.7,0.2\} = 0.2$$
$$\mu(P \cap Q)(y2) = \min\{\mu P(y2), \mu Q(y2)\} = \min\{0.3,0.5\} = 0.3$$
$$\mu(P \cap Q)(y3) = \min\{\mu P(y3), \mu Q(y3)\} = \min\{0.9,0.7\} = 0.7$$
$$\mu(P \cap Q)(y4) = \min\{\mu P(y4), \mu Q(y4)\} = \min\{0.1,0.4\} = 0.1$$



Gambar 10.4 Contoh operasi irisan pada himpunan fuzzy.

Representasi diagramatik administrator titik irisan ditunjukkan pada Gambar 10.4.

4. Asosiasi himpunan (Fuzzy) :

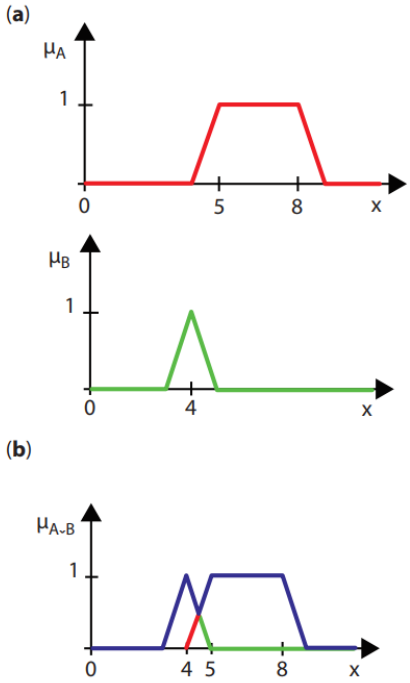
Asosiasi himpunan Fuzzy mencakup setiap komponen yang ditampilkan dalam himpunan. Estimasi nilai partisipasi akan menjadi estimasi partisipasi terbesar dari komponen di salah satu himpunan. Misalkan $P(y)$ dan $Q(y)$ adalah dua himpunan Fuzzy untuk semua $y \in X$.

Gabungan himpunan Fuzzy dimaksud dengan $(PUQ)(y)$ dan nilai kerja partisipasi diselesaikan sebagai berikut:

$$\mu(PUQ)(y) = \max\{\mu P(y), \mu Q(Y)\}$$

Contoh:

$$P(y) = \{(y1,0.7), (y2,0.3), (y3,0.9), (y4,0.1)\}$$
$$Q(y) = \{(y1,0.2), (y2,0.5), (y3,0.7), (y4,0.4)\}$$
$$\mu(PUQ)(y1) = \max\{\mu P(y1), \mu Q(y1)\} = \max\{0.7,0.2\} = 0.7$$
$$\mu(PUQ)(y2) = \max\{\mu P(y2)\mu Q(y2)\} = \max\{0.3,0.5\} = 0.5$$



Gambar 10.5 Contoh operasi gabungan pada himpunan fuzzy.

$$\mu(PUQ)(y3) = \max\{\mu P(y3), \mu Q(y3)\} = \max\{0.9,0.7\} = 0.9$$
$$\mu(PUQ)(y4) = \max\{\mu P(y4), \mu Q(y4)\} = \max\{0.1,0.4\} = 0.4$$

Catatan: Union sangat mirip dengan aktivitas OR yang sah yang ditunjukkan pada Gambar 10.5.

5. Hasil logaritma himpunan (Fuzzy) :

Hasil dari dua himpunan Fuzzy P(y) dan Q(y) untuk semua y ∈ Y, disebut P(y).Q(y) Dikarakterisasikan seperti yang diberikan di bawah ini:

$$P(y).Q(y) = \{(y, \mu P(y).\mu Q(y)), y \in Y\}$$

Contoh:

$$P(y) = \{(y1,0.1), (y2,0.2), (y3,0.3), (y4,0.4)\}$$
$$Q(y) = \{(y1,0.5), (y2,0.7), (y3,0.8), (y4,0.9)\}$$
$$P(y).Q(y') = \{(y1,0.05), (y2,0.14), (y3,0.24), (y4,0.36)\}$$

6. Hasil Perkalian Bilangan (Fuzzy) dengan Bilangan (Crisp) :
Hasil perkalian himpunan Fuzzy P(y) dengan Bilangan Crisp 'd' dirumuskan sebagai berikut.

$$P(y).Q(y) = \{(y, d.\mu P(y)), y \epsilon Y\}$$

Contoh:
Mari kita perhatikan himpunan Fuzzy $P(y)$ sedemikian rupa sehingga

$$P(y) = \{(y1,0.1), (y2,0.2), (y3,0.3), (y4,0.4)\}\ d = 0.2$$

pada titik tersebut $d.P(y) = \{(y1,0.02), (y2,0.04), (y3,0.06), (y4,0.08)\}$

7. Intensitas suatu himpunan (Fuzzy) :
Intensitas ke-p dari himpunan Fuzzy $P(y)$ menghasilkan himpunan Fuzzy lain $PP(y)$, yang nilai partisipasinya dapat diselesaikan sebagai berikut

$$\mu P^P(y) = \{\mu P(y)\}^P, y \epsilon y\}$$
$$p >= 1, P^P(y) \text{ called (fixtion)}$$
$$P < 1, P^P(y) \text{ is called (enlargement)}$$

Contoh:
Pertimbangkan himpunan Fuzzy $A(y)$

$$P(y) = \{y1,0.1), (y2,0.2), (y3,0.3), (y4,0.4)\}\ P = 2$$

Pada titik tersebut $P^2(y) = \{(y1, 0,01), (y2, 0,04), (y3, 0,09), (y4, 0,16)\}$

8. Bilangan Bulat Matematis Dua Himpunan (Fuzzy) :
Pada Fungsi ini, bilangan bulat dua himpunan Fuzzy P(y) dan Q(y) untuk semua y ∈ y, dinyatakan dengan P(y)+Q(y) dan dikarakterisasikan sebagai berikut

$$P(y) = \{Q(y) = \{y, \mu P + Q(y), y \epsilon Y\}$$

Di sini $\mu P + Q(y) = \mu P(y) + \mu Q(y) - \mu P(y).\mu Q(y)$ Contoh:

$$P(y) = \{(y1,0.5), (y2,0.2), (y3,0.3), (y4,0.4)\}$$

$$Q(y) = (y1,0.5), (y2,0.7), (y3,0.8), (y4,0.9)\}$$
$$\text{Presently } (y) + Q(y) = \{(y1,0.55), (y2,0.76), (y3,0.86), (y4,0.94)\}$$

10.3 PERENCANAAN SISTEM PAKAR BERBASIS ATURAN UNTUK KEAMANAN SIBER

Tingkat perencanaan menggabungkan karakterisasi faktor Sistem Pakar keamanan digital, pengumpulan informasi untuk Ancaman digital, struktur kerangka kerja dan penggunaan. Tingkat-tingkat ini digambarkan dalam bagian-bagian berikut.

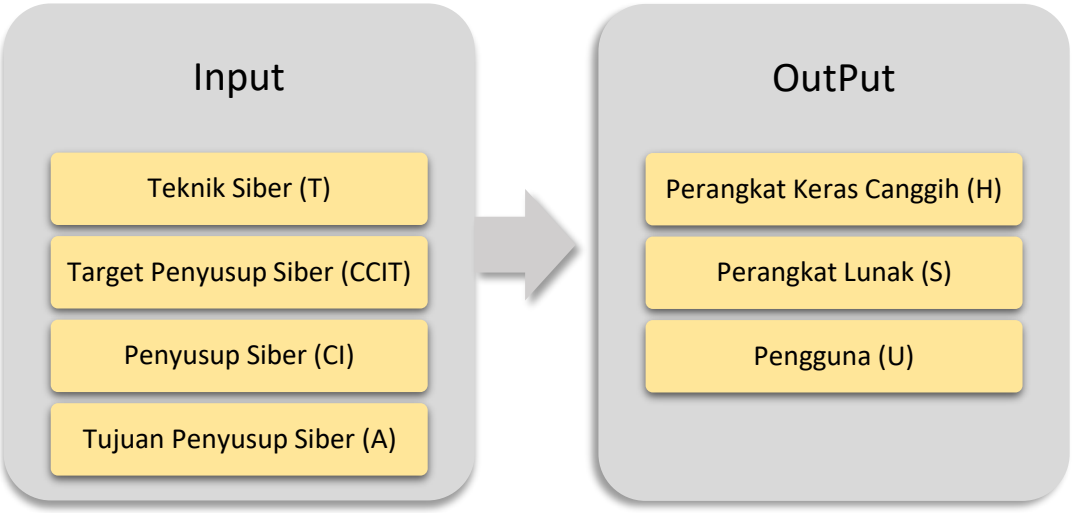
Tingkat 1: Menentukan Variabel Sistem Pakar Keamanan Siber

Fase awal dalam model yang diusulkan adalah fondasi faktor info dan hasil. Pekerjaan ini biasanya dilakukan dengan mempertimbangkan ruang yang sulit dan dengan konsultasi dengan spesialis digital. Ada sejumlah besar pelamar potensial yang harus dibatasi pada angka-angka positif.

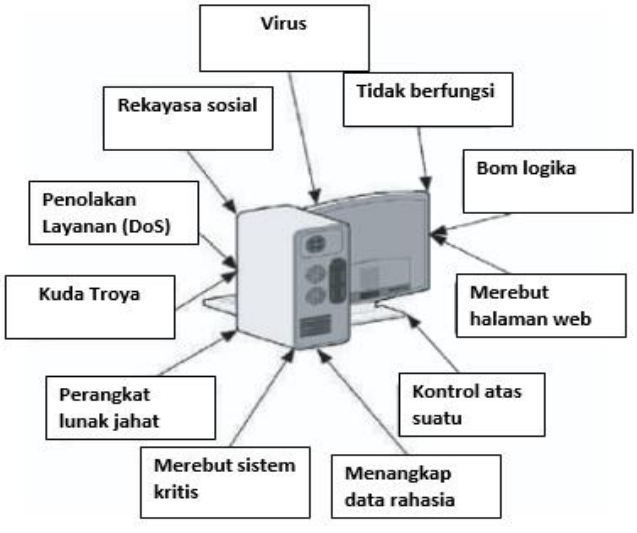
Tingkat 2: Pengumpulan Informasi untuk Terorisme Siber

Sistem Pakar menyusun data pada sistem manusia. Ini memberikan klarifikasi seperti sistem manusia. Informasi yang digunakan untuk pekerjaan ini telah diekstraksi dari serangkaian jajak pendapat yang dikumpulkan dari spesialis digital dan manajer kerangka kerja. Informasi yang tersimpan terhubung khususnya dengan poin-poin yang diberikan di bawah Gambar 10.6.

- virus, malware, bom logika, serangan DoS, rekayasa sosial.
- Kurangnya layanan, penyitaan halaman web, serangan untuk protes, penyitaan sistem kritis, penangkapan informasi rahasia.



Gambar 10.6 Model yang diusulkan untuk input dan output.



Gambar 10.7 Potensi ancaman siber.

Penganalisis menghitung sistem mana transportasi, pusat keuangan, sistem tenaga, layanan cepat, pasokan air, stasiun distribusi minyak dan gas alam yang mungkin diserang oleh teroris siber, seperti yang dijelaskan dalam Gambar 10.7.

Level 3: Desain Sistem

Sistem berbasis pengetahuan mungkin berlabuh maju atau mundur. Dalam kerangka kerja asosiasi maju, kita bernalar dari kebenaran pendahulu ke kebenaran yang dihasilkan; kita bernalar dari realitas dalam pendahulu standar yang kita pahami konsisten dengan pengembangan realitas baru yang kebenarannya disimpulkan oleh pendahulu. Dalam ikatan terbalik, ini terbalik; kita berupaya mencari kemampuan untuk membangun realitas beberapa keadaan opsional.

- Forward Chaining: Aturan kerangka kerja spesialis mungkin direncanakan pada dasarnya sebagai "jika Xn, maka Y" di mana X akan menjadi banyak keadaan pada informasi dan Y adalah banyak arah yang harus diselesaikan saat standar dimulai. Standar dianalisis untuk melihat aturan mana yang dibuat awal dapat dilakukan oleh informasi, yaitu, A terpenuhi, dan standar atau aturan dipilih untuk dijalankan. Pada saat standar dijalankan, susunan arahan Y dijalankan.
- Aturan terbalik: Pengelompokan alternatif diikuti dalam penjangkaran terbalik. Dalam penjangkaran terbalik, kami menghitung akhir yang mungkin ingin kami capai, yaitu, kami menunjukkan Y. Kami menemukan standar atau keputusan yang memiliki sub-sequent ideal, dan melihat pendahulu X untuk memahami informasi apa yang harus ada untuk memenuhi P. Sekarang kami menemukan bagaimana informasi itu dapat dibangun, dan mencari keputusan yang memiliki informasi itu sebagai sub-sequent, atau informasi informasi dari klien untuk memeriksa apakah pendahulu dapat dipenuhi. Dalam penjangkaran terbalik, kami bekerja secara terbalik dari tujuan ke informasi; dalam penjangkaran maju, kami bekerja maju dari informasi ke tujuan.

Tiga segmen fundamental diberikan di bawah ini:

- Antarmuka manusia.
- Motor penduga pengambilan keputusan.
- Kumpulan data (menyimpan informasi dan prinsip Fuzzy).

Master digital dapat terhubung dengan bantuan antarmuka Sistem berbasis Pengetahuan untuk menanyakan dan membaca panduan dari sistem baru. Motor derivasi terdiri dari Ancaman informasi digital, profil penyebar ketakutan digital, dan metode serangan digital.

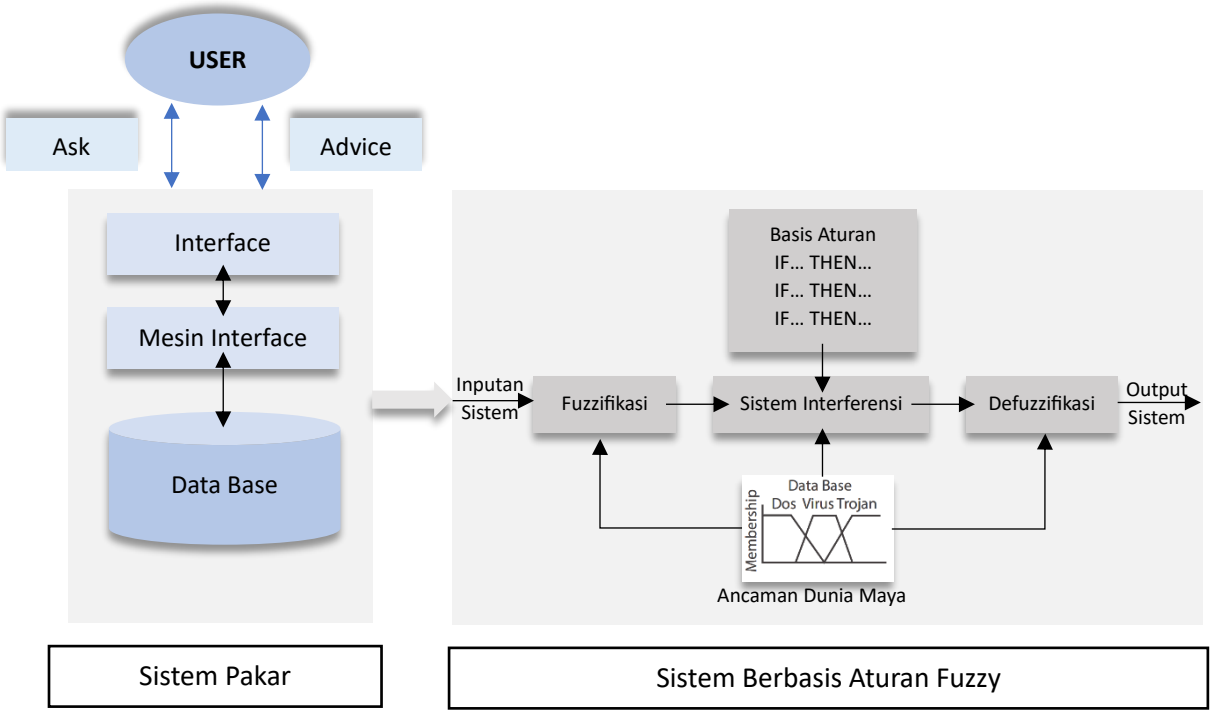
Level 4: Model Berbasis Aturan

Dalam model ini, rekayasa keseluruhan untuk sistem berbasis pengetahuan dan segmen kerangka kerja derivasi berbasis pedoman Fuzzy. Objek prinsip kerangka kerja berbasis standar Fuzzy adalah modul fuzzifikasi – atau fuzzifier – pedoman Fuzzy, motor induksi, dan defuzzifier.

Level 1. Model fuzzifikasi: Dalam fuzzifikasi, kontribusi ruang area objek informasi untuk dihitung dengan set (Fuzzy). Mengembangkan kapasitas partisipasi rasional Fuzzy mengasumsikan pekerjaan penting untuk struktur berbasis prinsip Fuzzy. Pekerjaan partisipasi tiga sisi digunakan dalam banyak aplikasi berbasis logika Fuzzy. Dalam pemeriksaan ini kapasitas partisipasi tiga sisi telah digunakan pada Gambar 10.8.

Level 2. Mengkarakterisasi pedoman Fuzzy: Standar Fuzzy terdiri dari prekursor dan hasil sebagai artikulasi IF-THEN. Ada berbagai prinsip, dan mereka membuat kumpulan yang membentuk alasan untuk deduksi.

Level 3. Defuzzifikasi: Koordinasi antara kontrol logika Fuzzy dan kerangka induksi, dengan memberikan hasil baru. Teknik defuzzifikasi umum adalah centroid, bisector, estimasi rata-rata kualitas terbesar, estimasi terkecil kualitas paling ekstrem, dan estimasi terbesar kualitas terbesar. Transformasi himpunan Fuzzy menjadi nilai baru soliter disebut defuzzifikasi dan prosedur yang berlawanan adalah fuzzifikasi.



Gambar 10.8 Model sistem berbasis aturan.

Ada sejumlah nilai paling ekstrem yang dapat diantisipasi dari strategi centroid, pengumpulan dipertahankan dalam pembuatan pengumpulan premis hambatan nilai terkecil dan terbesar dicirikan dalam instrumen bisector yang digunakan dalam model m fuzzifikasi berbasis aturan, nilai baru dan nilai Fuzzy dicirikan pada nilai numerik. Nilai biasa diperlukan dari antarmuka untuk menunjukkan keluaran baru.

10.4 KEAMANAN DIGITAL

Ancaman Siber

Kemajuan berbasis digital saat ini tersebar luas di seluruh dunia. Sejauh ini, sebagian besar klien mencari tujuan yang sah, cakap, dan individual. Bagaimanapun, pelanggar hukum, penindas psikologis, dan mata-mata juga sangat bergantung pada kemajuan berbasis digital untuk membantu mereka mencapai target. Pelanggar ini dapat mengakses kemajuan berbasis digital untuk menolak bantuan, mengambil atau mengendalikan informasi, atau menggunakan gadget untuk meluncurkan serangan terhadap dirinya sendiri atau perangkat lain. Elemen yang menggunakan kemajuan berbasis digital untuk tujuan ilegal mengambil banyak bentuk.

Kesalahan Cyber

Kesalahan sistem, kesalahan digital, kesalahan elektronik, kesalahan elektronik sebagian besar termasuk kejahatan di mana suatu sistem adalah sumber, perangkat, target, atau tempat kesalahan. Dalam kesalahan ini, pengelompokannya tidak selektif dan banyak kegiatan dapat digambarkan sebagai termasuk dalam setidaknya satu kelas. Selain itu, meskipun istilah kesalahan sistem atau kejahatan dunia maya lebih tepat dibatasi untuk menggambarkan kejahatan di mana sistem atau sistem merupakan bagian penting dari kesalahan, istilah-istilah ini juga terkadang digunakan untuk memasukkan pelanggaran konvensional, misalnya, penipuan, pencurian, pemaksaan, pemalsuan, dan pencurian, di mana sistem atau sistem digunakan untuk mendorong pergerakan yang melanggar hukum.

Kesalahan digital juga merupakan masalah yang signifikan saat ini secara global; banyak orang yang meretas kerangka sistem. Kesalahan sistem secara komprehensif dapat dicirikan sebagai kejahatan yang mencakup kerangka inovasi data, termasuk akses ilegal (akses tidak sah), upaya pemblokiran ilegal (dengan metode khusus untuk transmisi informasi sistem yang tidak terbuka ke, dari atau di dalam kerangka sistem), penghalangan informasi (perusakan tidak sah, pembatalan, disintegrasi, modifikasi atau penyembunyian informasi sistem), hambatan sistem (campur tangan dengan kerja kerangka sistem dengan berkontribusi, mengirim, merusak, menghapus, menghancurkan, memodifikasi atau mencekik informasi PC), penyalahgunaan gadget, peniruan (perampokan ID), dan pemerasan elektronik. Berikut adalah jenis-jenis kesalahan sistem:

a. Virus komputasi
b. Aktivitas penipuan
c. Kode Berbahaya
d. Serangan Penolakan Layanan
e. Peretasan
f. Kejahatan (cyber)
g. Terorisme Cyber
h. Perang Informasi
i. Cyber Stalking
j. Penipuan dan Pencurian Identitas
k. Kejahatan (virtual)

Berbagai Jenis Layanan Keamanan

Mekanisme keamanan merupakan komponen penting yang terlibat dengan suatu sistem. Saat klien berbagi aset dan informasi pada suatu sistem, mereka harus memiliki opsi untuk mengontrol siapa yang dapat mengakses informasi atau aset tersebut dan apa yang dapat dilakukan klien dengannya. Contohnya adalah dokumen yang menunjukkan catatan keuangan suatu organisasi. Jika dokumen ini ada pada pekerja catatan, penting untuk memiliki opsi untuk mengontrol siapa yang mengakses dokumen tersebut. Selain itu, siapa yang dapat membaca dan mengubah catatan juga merupakan pemikiran penting. Model yang sama ini juga berlaku untuk printer umum. Anda harus menentukan siapa yang dapat menggunakan printer laser warna mahal atau, lebih khusus lagi, kapan seseorang dapat menggunakan printer ini.

Seperti yang seharusnya jelas, keamanan merupakan bantuan yang signifikan pada suatu sistem. Manajer sistem menghabiskan banyak waktu untuk mempelajari dan menyiapkan keamanan. Manfaat keamanan sering kali mengelola basis data akun klien atau sesuatu seperti administrasi indeks yang disebutkan sebelumnya. Basis data klien sering kali berisi daftar nama pengguna dan nomor acak. Seseorang perlu mengakses sistem; ia harus masuk ke sistem. Masuk seperti mencoba memasuki tempat usaha dengan petugas keamanan di pintu masuk depan. Sebelum Anda dapat memasuki bangunan, Anda harus memeriksa siapa Anda terhadap daftar orang-orang yang diizinkan masuk. Manfaat keamanan sering kali dicampur dengan berbagai layanan. Beberapa layanan yang ditambahkan ke sistem dapat menggunakan layanan keamanan dari kerangka kerja tempat mereka telah dipasang.

Peningkatan Sistem Keamanan Siber

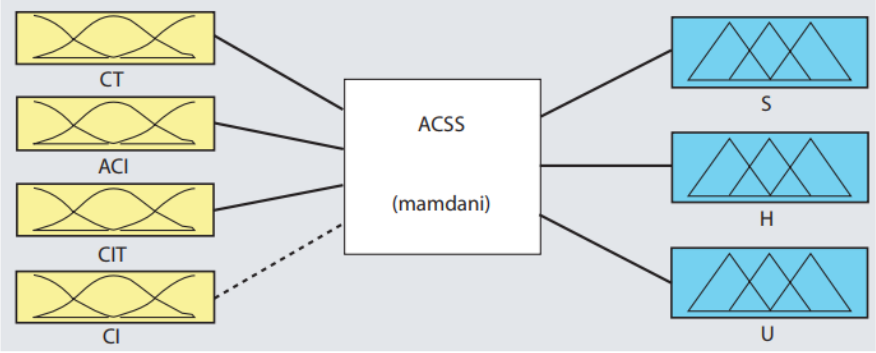
Pada titik ini, tulisan terkini tentang kerangka kerja keamanan digital telah diringkas, dan batasan dasar dari masa lalu ditampilkan. Tingkat penataan mencakup penggambaran faktor kerangka kerja keamanan digital, pengumpulan informasi untuk Ancaman digital, rencana dan penggunaan kerangka kerja. Tingkat-tingkat tersebut didefinisikan di bawah ini:

Struktur

Tingkat awal dalam sistem baru adalah dasar dari faktor informasi dan hasil. Tugas ini biasanya dilakukan dengan memeriksa ruang yang sulit. Terdapat jumlah calon pendatang baru yang tak ada habisnya yang seharusnya dibatasi pada angka positif yang dapat dijelaskan pada Gambar 10.9.

Variabel Input	Singkatan	Variabel Keluaran	Singkatan
Teknik Cyber	TC	Perangkat lunak	PL
Tujuan Penyusup Cyber	TPC	Perangkat keras	PK
Target Penyusup Siber	TPS	Pengguna	P
Penyusup Siber	PS		

Gambar 10.9 Variabel input output.



Gambar 10.10 Struktur sistem keamanan siber.

Terorisme Siber untuk Pengumpulan Informasi/Data

Pada bagian ini, struktur pengembangan, informasi pada master manusia. Ini mendefinisikan klarifikasi seperti pakar manusia. Model dapat merancang analisis khas oleh pengguna. Informasi yang digunakan untuk pekerjaan ini telah dihapus dari serangkaian survei yang dikumpulkan dari eksekutif kerangka kerja digital yang dijelaskan pada Gambar 10.10. Poin-poin pentingnya adalah:

- Serangan DoS, infeksi, bom alasan, perancangan sosial.
- Keluar dari administrasi, memegang halaman situs, serangan untuk pembangkangan, memegang kerangka kerja dasar, menangkap data rahasia, kontrol kerangka kerja.

Investigasi ini menilai penindas berbasis ketakutan digital yang dapat menyerang kerangka kerja korespondensi, pusat keuangan, pembangkit listrik, administrasi krisis, transportasi, pasokan air, stasiun alokasi minyak dan gas bumi. Individu yang diperlengkapi untuk penindasan psikologis digital, misalnya, staf khusus yang berdedikasi, programmer, dan sistem model digital.

Kesimpulan

Kerangka kerja spesialis untuk keamanan digital bergantung pada prinsip Fuzzy. Setelah bertemu dengan spesialis digital dan direktur kerangka kerja, sumber data dan hasil kerangka kerja diselesaikan. Pengurangan standar Fuzzy diselesaikan dengan menggunakan administrator 'min' dan 'maks' untuk konvergensi dan asosiasi Fuzzy. Ruang informasi diisolasi ke dalam alokasi multidimensi untuk menentukan basis pedoman yang mendasarinya.

Aktivitas kemudian diturunkan ke setiap segmen. Investigasi ini mengusulkan penanda digital berbasis prinsip Fuzzy yang memperingatkan direktur kerangka kerja untuk aktivitas digital. Ditemukan bahwa kerangka kerja berfungsi dengan baik yang kondisinya cocok dengan aktivitas digital. Dorongan beberapa sinyal peringatan dibuat adil dan jujur. Tujuan model bukanlah untuk mengamankan kerangka kerja tetapi menargetkan peringatan ketua kerangka kerja untuk aktivitas digital yang diharapkan. Kerangka kerja keamanan digital pengembangan yang bergantung pada pedoman Fuzzy diperkenalkan. Kerangka kerja induksi Fuzzy dipilih untuk membuat sistem keamanan siber.

Pengurangan standar Fuzzy diselesaikan dengan menggunakan administrator 'minimum' dan 'maksimum' untuk titik persimpangan dan asosiasi Fuzzy. Ruang info dipisahkan menjadi beberapa paket multidimensi untuk merencanakan basis prinsip yang mendasarinya. Aktivitas kemudian dibagikan ke setiap segmen. Memastikan bahwa korespondensi informasi melalui web dan beberapa sistem lainnya secara konsisten berada di bawah Ancaman gangguan dan penyalahgunaan. Jadi Sistem Deteksi Intrusi telah menjadi segmen yang dibutuhkan sejauh menyangkut keamanan PC dan sistem.

Ada berbagai metodologi yang digunakan di lokasi gangguan; namun, tidak ada kerangka kerja sejauh ini yang benar-benar sempurna. Sejalan dengan itu, misi perbaikan berlanjut. Strategi penalaran fuzzy memberikan pendekatan untuk menggambarkan faktor-faktor yang dicirikan secara longgar, mengkarakterisasi hubungan antara faktor-faktor yang bergantung pada informasi manusia utama dan menggunakannya untuk memproses hasil. Sistem Pakar Fuzzy yang diterapkan pada bidang keamanan data adalah metode yang memadai untuk menyalin kapasitas dinamis profesional.

BAB 11
ANALISIS ANCAMAN MENGGUNAKAN TEKNIK PENAMBANGAN DATA

Dengan kemajuan teknologi informasi, penggunaan internet memainkan peran penting dalam aktivitas sehari-hari, dan karena itu terorisme siber meningkat pesat. Penjahat siber melakukan banyak serangan siber seperti Phishing, Denial of service, serangan Kata sandi, dll. Karena kurangnya metode komputasi, pendekatan teknis yang ada tidak cukup untuk menyelidiki dan mengendalikan serangan siber. Oleh karena itu, skenario saat ini memerlukan pendekatan yang lebih maju untuk memperbaiki masalah serangan siber. Tujuan dari bab ini adalah untuk menganalisis ancaman siber dan untuk menunjukkan bagaimana pendekatan kecerdasan buatan dan penambangan data dapat efektif untuk memperbaiki masalah serangan siber.

Bidang kecerdasan buatan telah memainkan peran yang semakin penting dalam menganalisis ancaman siber dan meningkatkan keamanan serta keselamatan siber. Tiga aspek utama dibahas dalam bab ini. Pertama, proses deteksi serangan siber yang akan membantu menganalisis dan mengklasifikasikan insiden siber. Kedua, peramalan serangan siber yang akan datang dan pengendalian terorisme siber. Terakhir, bab ini berfokus pada latar belakang teoritis dan kegunaan praktis kecerdasan buatan dengan pendekatan penambangan data untuk mengatasi masalah di atas melalui deteksi dan prediksi.

11.1 PENDAHULUAN

Dalam beberapa tahun terakhir, seluruh dunia telah bergantung pada teknologi informasi dan komunikasi TIK untuk pekerjaan profesional, hiburan, pendidikan, dan kehidupan sosial. Alasan di balik ketergantungan ini adalah popularitas IoT (Internet of Things), perluasan jaringan komputer yang luar biasa, dan banyaknya aplikasi yang digunakan oleh berbagai kelompok serta individu untuk penggunaan pribadi dan profesional mereka. Permintaan keamanan siber telah meningkat karena berbagai serangan siber seperti akses tidak sah, serangan penolakan layanan, malware komputer, dll. Sistem keamanan komputer dan jaringan menghasilkan keamanan siber.

Sistem enkripsi dan firewall tersedia untuk mengelola serangan siber. Saat ini, solusi konvensional seperti firewall tidak dapat menjalankan tugasnya. Karena kurangnya metode komputasi, pendekatan teknis yang ada tidak cukup untuk menyelidiki dan mengendalikan serangan siber. Oleh karena itu, skenario saat ini memerlukan pendekatan yang lebih maju untuk memperbaiki masalah serangan siber di atas. Bab ini menyajikan metode pembelajaran mesin (ML) dan penambangan data (DM) untuk keamanan siber. Tujuan dari bab ini adalah untuk menganalisis ancaman siber dan menunjukkan bagaimana pendekatan kecerdasan buatan dan penambangan data dapat efektif untuk memperbaiki masalah serangan siber.

Bidang kecerdasan buatan telah memainkan peran yang semakin penting dalam menganalisis ancaman siber dan meningkatkan keamanan siber serta keselamatan. Saat ini Kecerdasan Buatan adalah alat yang mengubah dunia, yang meningkatkan kemampuan manusia dalam berbagai bidang. Tiga aspek utama dibahas dalam bab ini. Pertama, proses

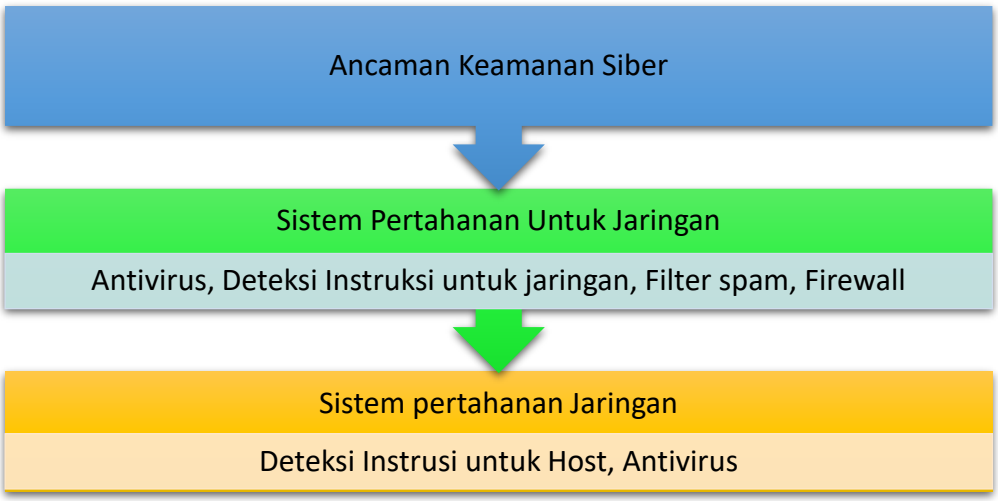
deteksi serangan siber, yang akan membantu menganalisis dan mengklasifikasikan insiden siber. Kedua, meramalkan serangan siber yang akan datang dan mengendalikan terorisme siber. Terakhir, bab ini berfokus pada latar belakang teoritis dan kegunaan praktis kecerdasan buatan dengan pendekatan penambangan data untuk mengatasi masalah di atas melalui deteksi dan prediksi.

Setelah pendahuluan bagian 11.1, bagian yang tersisa disusun dengan cara berikut. Bagian 11.2 memberikan latar belakang dan pekerjaan terkait serangan siber dan keamanan siber dengan teknik penambangan data (Kecerdasan buatan dan pembelajaran mesin). Di Bagian 11.3, berbagai metode berbasis Penambangan Data dibahas untuk mendeteksi serangan siber. Bagian 11.4 menjelaskan proses deteksi serangan siber berdasarkan penambangan data. Terakhir, kesimpulan bab ini ada di Bagian 11.5, yang juga menyoroti cakupan di masa mendatang.

11.2 KEAMANAN TEKNOLOGI INFORMASI TI

Keamanan Siber

Disebut juga sebagai keamanan teknologi informasi TI. Keamanan siber adalah salah satu jenis badan teknologi dan proses yang melindungi jaringan komputer, perangkat lunak, dan perangkat komputer dari berbagai jenis kerusakan. Saat ini keamanan siber diminati oleh berbagai sektor pemerintah, Pertahanan dan Angkatan Darat, lembaga keuangan, organisasi medis dan kesehatan. Tujuannya adalah untuk menjaga kerahasiaan, kebenaran, serta ketersediaan sistem manajemen informasi oleh sistem pertahanan siber. Para peneliti dari akademisi, sektor pemerintah, dan industri swasta telah terlibat dalam merancang dan melaksanakan berbagai jenis sistem pertahanan siber (seperti yang dijelaskan dalam Gambar 11.1).



Gambar 11.1 Sistem konvensional untuk keamanan siber.

Lanskap Ancaman Siber

Meskipun kategori ancaman umum tidak berubah, lanskap musuh dan pembuat onar dalamkeamanan komputer telah berkembang seiring waktu. Untuk tujuan tersebut, penting untuk memiliki pengetahuan yang baik tentang berbagai kemungkinan serangan.

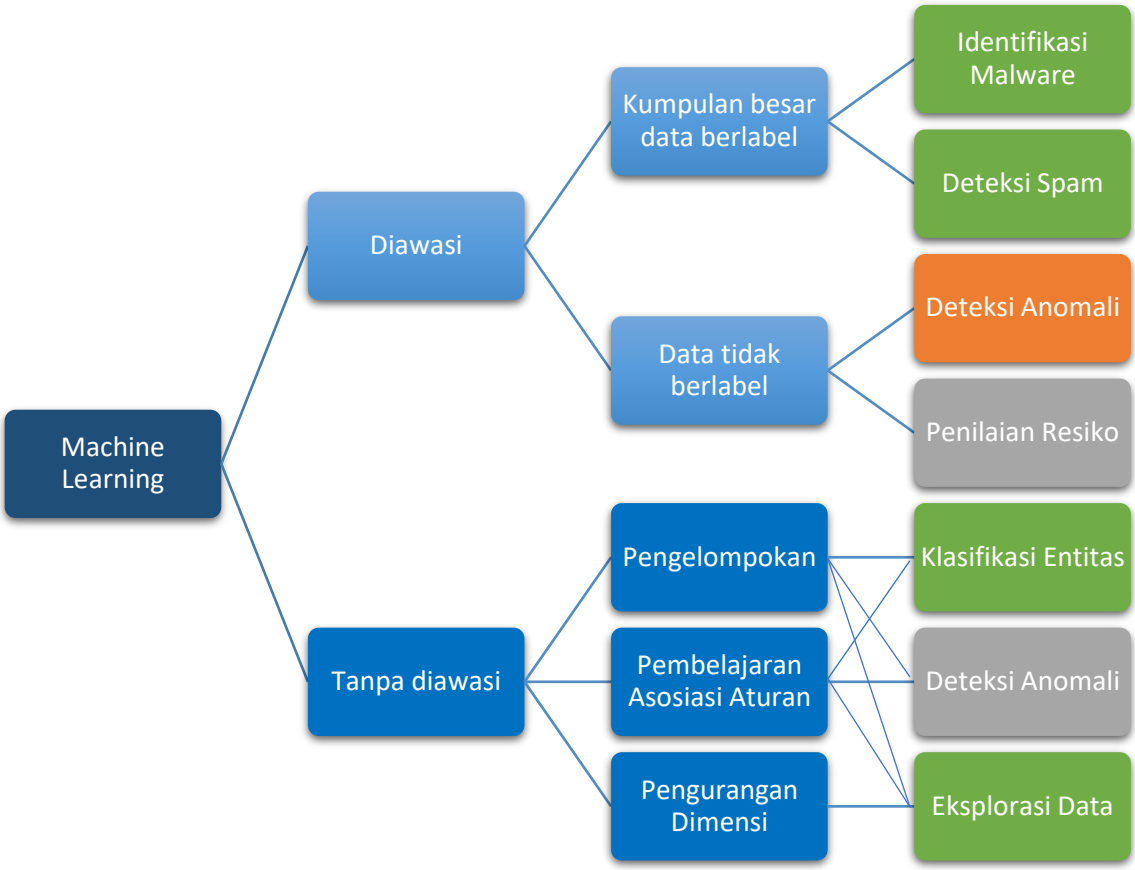
Tabel 11.1 Serangan siber umum dan deskripsinya.

Serangan siber	Description
Malware/Trojan Virus	Perangkat lunak yang sengaja dirancang untuk menyebabkan kerusakan pada sistem komputer
Spyware	Malware yang dipasang pada sistem komputer tanpa izin untuk tujuan infiltrasi dan pengumpulan informasi
Adware	Malware yang didukung iklan
Rootkit	Menyebabkan kekacauan di latar belakang komputer
backdoor	Lubang yang direncanakan ditempatkan memungkinkan akses di masa mendatang tanpa perlindungan perimeter.
Bot	Malware yang menyebar sendiri
Botnet	Jaringan bot yang besar.
Exploit	Kode yang diuntungkan dari kerentanan perangkat lunak
Scanning	Dengan mengirimkan permintaan ke sistem komputer secara bruteforce untuk mencari titik lemah, pengumpulan informasi, dan kerentanan.
Sniffing	Virus mata-mata yang memungkinkan untuk merekam informasi pribadi
Keylogger	Rekam informasi penekanan tombol dengan perangkat lunak/perangkat keras
Spam	Menyebar dari satu sistem ke sistem lain melalui email
ATO	Pengambilalihan akun; Mendapatkan akses ke akun orang lain
Phishing	Penyerang menampilkan diri sebagai bisnis legal untuk menipu dan mendapatkan informasi pribadi seseorang
DoS	Serangan penolakan layanan melalui pemboman volume tinggi dan/atau permintaan yang salah format

Kecerdasan Buatan (AI) dan Pembelajaran Mesin (ML)

AI merupakan upaya individu untuk membuat mesin lebih cerdas. Definisi AI sedikit lebih kontroversial daripada definisi pembelajaran mesin. Kecerdasan Buatan didefinisikan sebagai keputusan yang berorientasi pada mesin untuk menyelesaikan tugas-tugas otak tingkat manusia, dan pembelajaran mesin didefinisikan sebagai pembelajaran dari data masa lalu untuk memprediksi masa depan. Ada dua jenis utama pembelajaran mesin:

1. Pembelajaran mesin yang diawasi: Data berlabel memberikan umpan balik langsung dan memprediksi masa depan.
2. Pembelajaran mesin tanpa pengawasan: Tidak ada label, tidak ada umpan balik, hanya menemukan struktur tersembunyi dalam data.



Gambar 11.2 Keamanan siber dengan pembelajaran mesin yang diawasi dan tanpa pengawasan.

11.3 METODE PENAMBANGAN DATA YANG MENDUKUNG DETEKSI SERANGAN SIBER

Bagian saat ini menggambarkan metode Penambangan Data yang mendukung serangan siber dan terkait dengan keamanannya. Teknik penambangan data seperti aturan asosiasi, klasifikasi, pengelompokan umumnya digunakan untuk mendeteksi berbagai jenis serangan siber.

Klasifikasi

Klasifikasi, juga disebut pengklasifikasi, yang menetapkan objek data ke kelas yang telah ditentukan sebelumnya. Pertama-tama, dengan menganalisis satu set pelatihan, pengklasifikasi dilatih. Di sini set pelatihan ditemukan dari contoh data dan label kelas terkaitnya. Jenis pendekatan ini adalah pembelajaran terbimbing karena label kelas contoh pelatihan tersedia. Kedua, dengan menggunakan pengklasifikasi pelatihan, proses peramalan kelas untuk contoh data yang tidak berlabel dilakukan. Semua kelas telah ditentukan sebelumnya dalam fase pelatihan. Umumnya klasifikasi memiliki dua jenis kasus. Klasifikasi Biner: Di sini hanya dua kelas yang terlibat.

A. Klasifikasi Multikelas: Di sini beberapa kelas terlibat.

Untuk penyalahgunaan serta deteksi anomali, klasifikasi dapat digunakan. Dalam data audit, setiap contoh data diberi label sebagai data Normal atau data Abnormal. Algoritma klasifikasi diterapkan pada data audit untuk tujuan melatih pengklasifikasi,

yang akan digunakan untuk tujuan prediksi apakah data contoh baru akan menjadi "normal" atau "abnormal". Beberapa metode klasifikasi yang populer adalah K-nearest neighbor, pengklasifikasi Naive Bayes, pohon keputusan, support vector machine, logika fuzzy, Jaringan Syaraf Tiruan.

Pohon Keputusan:

Ini adalah jenis formasi pohon dengan daun. Pohon keputusan melambangkan klasifikasi di mana cabang-cabangnya melambangkan penyatuan fitur yang akan mengarah pada klasifikasi. Aturan "IF THEN" adalah pilar dari pohon keputusan. Struktur yang paling sederhana serta dapat ditafsirkan memungkinkan pohon keputusan untuk memecahkan masalah seperti atribut multi-tipe. Pohon keputusan digunakan untuk mengelola data gangguan atau nilai yang hilang. Pohon keputusan menyediakan implementasi yang sangat sederhana dan mudah.

K-Nearest Neighbor (KNN):

Algoritma ini pada dasarnya adalah algoritma klasifikasi, yang sangat mudah. Semua kasus saat ini yang merupakan data pelatihan, kasus baru yang terakumulasi dan yang akan datang yang merupakan data uji, akan diklasifikasikan tergantung pada ukuran kesamaannya dari ruang fitur yang diberikan. Jarak data instans baru dari kasus baru dan kasus yang ada dihitung kemudian data uji dipilih ke kelas yang akan lebih familiar di antara KNN (K-nearest neighbor)-nya. Di sini, jika nilai k=1, Maka dialokasikan ke kelas tetangga terdekatnya. Jika nilai k besar maka diperlukan waktu prediksi yang lebih lama.

Penambangan Aturan Asosiasi:

Ini akan menemukan hubungan antara berbagai variabel yang ada dalam basis data. Setelah integrasi data dan pembersihan data, penambangan aturan asosiasi akan diterapkan. Dalam penambangan aturan asosiasi, fase pertama untuk pembuatan set item permintaan kemudian tahap berikutnya pembuatan aturan, yang akan membuat aturan atribusi serangan siber.

Misalkan Jika (X dan Y) maka Z.

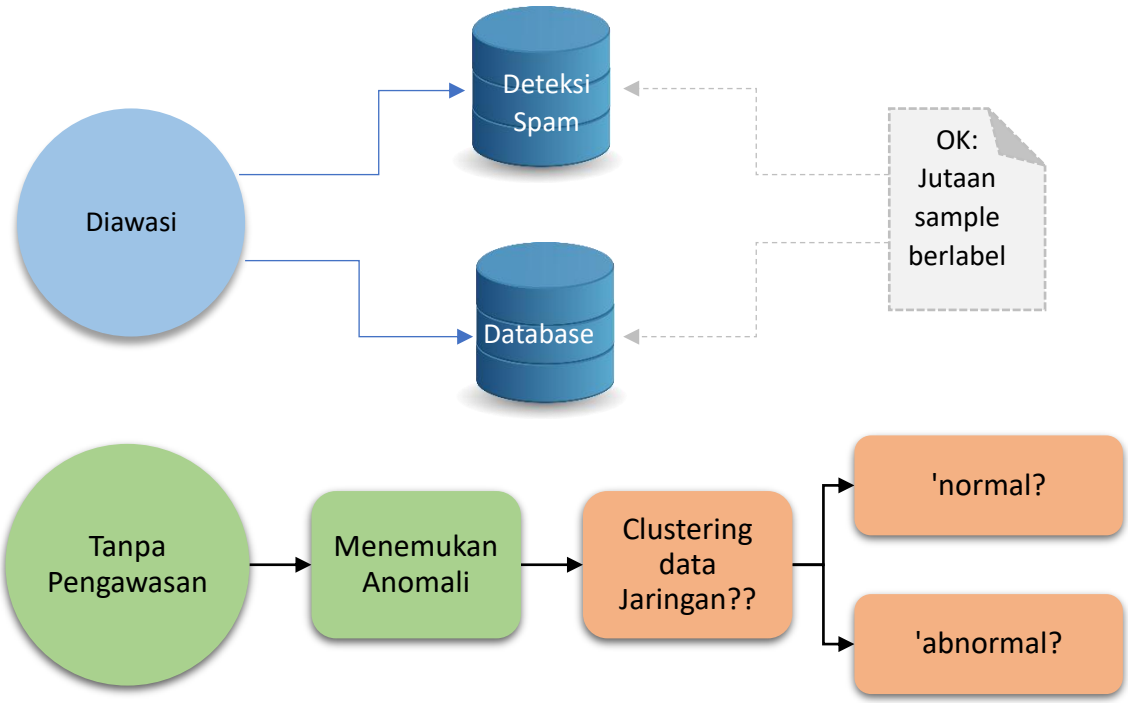
Berarti jika X dan Y keduanya tersedia maka z juga tersedia. Untuk deteksi serangan siber, aturan asosiasi yang sama diterapkan. Jika (Bukti1 DAN Bukti2) MAKA identifikasi kriminal. Gambar 11.3 menjelaskan cara membuat aturan atribusi serangan siber.

Pengelompokan :

Istilah pengelompokan berarti mendistribusikan berdasarkan beberapa kesamaan. Dalam deteksi intrusi, pengelompokan belajar dari data audit tanpa administrator sistem dengan memberikan rincian kelas serangan yang berbeda (ditunjukkan pada Gambar 11.4). Pengelompokan dikategorikan menjadi dua jenis, yaitu,

A. Pengelompokan Keras: item hanya dapat ditetapkan ke dalam satu kluster.
B. Pengelompokan Lunak: item dapat ditetapkan ke dalam beberapa kluster.

Gambar 11.3 Aturan atribusi serangan siber melalui penambangan aturan asosiasi.



Gambar 11.4 Pembelajaran terbimbing (Klasifikasi) dan Pembelajaran tak terbimbing (Pengelompokan).

11.4 PROSES DETEKSI SERANGAN SIBER BERDASARKAN PENAMBANGAN DATA

Penjelasan terperinci tentang proses Deteksi Serangan Siber berdasarkan Penambangan Data ditunjukkan pada Gambar 11.5. Pada tahap 1, Pemrosesan Data mencakup pemantauan sistem serta penangkapan data melalui berbagai sensor, pencatatan sistem atau jaringan serta agen pengendus atau daemon.

Alat penambangan data untuk keamanan siber:

Penambangan data memerlukan kumpulan data yang besar dan berbagai algoritme penambangan data diterapkan pada kumpulan data yang besar tersebut. Proses ini juga memerlukan beberapa analisis statistik, yang tidak mungkin dilakukan secara manual. Saat ini

banyak alat penambangan data yang tersedia; beberapa alat gratis. Alat penambangan data yang populer tercantum dalam Tabel 11.2.

Gambar 11.5 Berbagai tahap untuk mendeteksi serangan siber melalui penambangan data.

Tabel 11.2 Alat Penambangan Data Populer untuk keamanan siber.

Nomor sr.	Alat Penambangan Data
1.	Sistem Analisis Statistik Penambangan Data SAS
2.	Teradata
3.	Pemrograman R
4.	RapidMiner
5.	Oracle BI Business Intelligence
6.	KNIME
7.	Tanagra
8.	Weka Lingkungan Waikato untuk Analisis Pengetahuan
9.	Python
10.	Pemodel IBM SPSS

Penambangan data memiliki kekuatan yang cukup untuk mendeteksi malware. Penambangan data memungkinkan pemisahan sejumlah besar data dan memusatkan pembelajaran baru darinya. Penambangan data memiliki kemampuan untuk menemukan serangan yang diketahui maupun serangan zero-day. Bab ini membahas tentang analisis ancaman menggunakan berbagai metode, algoritma, dan alat penambangan data. Tahap penambangan data yang paling signifikan dalam analisis ancaman siber adalah kumpulan data untuk pelatihan dan pengujian. Di masa mendatang, Algoritma pembelajaran mendalam yang lebih canggih akan dianalisis untuk mendeteksi dan memprediksi serangan siber.

BAB 12
DETEKSI INTRUSI MENGGUNAKAN PENAMBANGAN DATA

Saat ini, internet merupakan metode komunikasi yang hampir universal bagi individu dan bisnis. Karena meningkatnya penggunaan internet, perspektif keamanannya menjadi semakin penting setiap hari untuk berbagai sistem deteksi intrusi jaringan (IDS) dari beberapa serangan. Beberapa IDS ditempatkan di lokasi jaringan yang heterogen untuk menjaganya. Berbagai metode digunakan untuk mendeteksi serangan atau penipuan dan dapat diterapkan dalam perspektif pohon keputusan. Ini memberikan cara paling sederhana untuk mengenali area yang paling tepat untuk memilih, mengelola, dan membentuk keputusan yang optimal mengenai identifikasi mereka dari kumpulan data terbesar.

Bab ini mengeksplorasi deteksi intrusi modern dengan perspektif penentuan yang khas dari penambangan data. Pembahasan ini berfokus pada aspek utama strategi deteksi intrusi, yaitu, deteksi penyalahgunaan. Ini berfokus pada identifikasi serangan, informasi atau data yang ada di jaringan menggunakan algoritma C4.5, yang merupakan jenis teknik pohon keputusan dan juga membantu meningkatkan sistem IDS untuk mengenali jenis serangan di jaringan. Untuk deteksi serangan ini, digunakan kumpulan data KDD-99; berisi beberapa fitur dan kelas data umum dan jenis serangan yang berbeda.

12.1 PENDAHULUAN

Akhir-akhir ini, setiap perusahaan dan lembaga menggunakan internet untuk pertukaran verbal serta, untuk media perusahaan komersial, untuk menjangkau pelanggan. Karena penggunaan internet semakin cepat, peningkatan penipuan komunitas juga meningkat, yang mana jaminan diri yang rendah dalam konektivitas struktur komunitas dan aset mereka telah memperluas potensi kerugian karena adanya aktivitas penipuan, yang diluncurkan ke sistem dari properti yang jauh. Sangat sulit untuk menyelamatkan semua orang dari pemerasan dengan metode untuk memanfaatkan firewall karena pada setiap kejadian, misrepresentasi eksplisit menggabungkan kekurangan atau bug yang tidak jelas.

Oleh karena itu, kerangka kerja pengenalan gangguan berkelanjutan digunakan untuk menemukan penipuan dan secara luas digunakan untuk menghentikan serangan yang sedang dikembangkan; ia menawarkan petunjuk peringatan kepada klien atau manajer jaringan yang disetujui tentang adanya minat jahat atau adanya penipuan. Tujuan IDS adalah untuk menemukan gangguan langsung ke PC atau organisasi, dengan memperhatikan berbagai olahraga atau karakteristik organisasi. Di sini gangguan mengacu pada setiap pengaturan perkembangan yang membahayakan kejujuran, aksesibilitas, atau klasifikasi aset bermanfaat organisasi.

Deteksi intrusi terdiri dari percabangan perangkat dan strategi yang mencakup penguasaan perangkat, catatan, penambangan statistik, dan sebagainya untuk identifikasi serangan. Dalam beberapa tahun terakhir, metode penambangan informasi untuk perangkat deteksi intrusi komunitas telah memberikan akurasi yang luar biasa dan deteksi yang tepat dari beberapa jenis penipuan. Teknik pohon pilihan adalah salah satu strategi klasifikasi intuitif dan

jujur dalam penambangan kebenaran yang dapat digunakan untuk tujuan ini. Teknik ini memiliki manfaat berkualitas tinggi dalam mengekstraksi fitur dan kebijakan.

Jadi, pohon pilihan memberikan signifikansi lebih besar pada deteksi intrusi. Pohon ini dibangun dengan sumber daya untuk mengidentifikasi atribut dan nilai terkaitnya dengan cara yang mengagumkan bagi pengguna untuk menguji statistik input di setiap simpul perantara pohon. Setelah pohon digunakan, pohon ini dapat merekomendasikan catatan yang baru datang dengan bantuan cara melintasi, awal dari simpul akar ke simpul daun dengan menggunakan perjalanan semua simpul internal dalam arah yang bergantung pada lingkungan pemeriksaan atribut di setiap simpul. Kerumitan nomor satu dalam membangun pohon pilihan adalah harga mana yang dipilih untuk membagi simpul pohon.

12.2 KONSEP DETEKSI INTRUSI

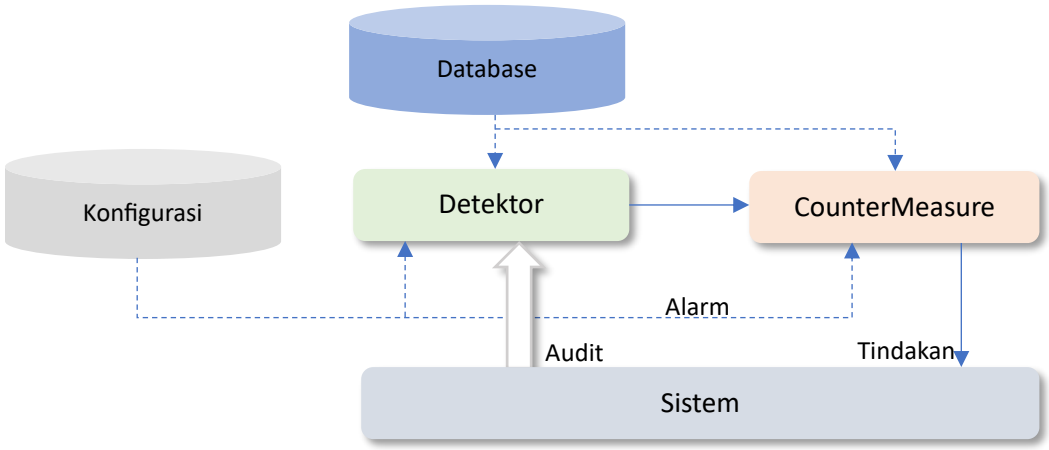
Dalam segmen ini, kita membahas IDS beserta strategi kelasnya dan tentang sejumlah kelas serangan. Segmen ini juga membahas secara khusus tentang bagaimana pohon pemilihan dibangun dan berbagai strategi dalam pohon keputusan untuk membuat pilihan yang tepat.

Sistem Deteksi Intrusi

Sistem deteksi intrusi (IDS) adalah mesin yang menampilkan lalu lintas jaringan unit untuk minat yang mencurigakan dan peringatan masalah saat minat tersebut ditentukan. Ini adalah utilitas perangkat lunak yang memindai jaringan atau sistem untuk aktivitas berbahaya atau pelanggaran cakupan. Pemberitahuan tentang setiap penugasan atau pelanggaran yang berbahaya biasanya dikirim ke administrator atau dikumpulkan secara terpusat menggunakan perangkat kontrol fakta dan peristiwa keamanan (SIEM).

Perangkat SIEM mengintegrasikan keluaran dari lebih dari satu sumber dan menggunakan teknik penyaringan alarm untuk membedakan minat berbahaya dari alarm palsu. Meskipun struktur deteksi intrusi mengungkapkan jaringan untuk aktivitas yang berpotensi berbahaya, mereka juga cenderung menghasilkan alarm palsu. Akibatnya, perusahaan ingin mempercepat produk IDS mereka setelah pertama kali menerapkannya, sehingga mereka dapat mengetahui seperti apa pengunjung situs biasa di komunitas tersebut dibandingkan dengan lobi jahat.

Meskipun struktur pengenalan gangguan mengungkap jaringan untuk kemungkinan tindakan ganas, struktur tersebut juga diatur untuk memalsukan peringatan. Dengan demikian, organisasi perlu melacak item IDS mereka setelah pertama kali mengirimkannya. Struktur penemuan gangguan akan menangkap seperti apa pengunjung situs biasa di jaringan tersebut jika dibandingkan dengan aktivitas ganas. Sesuai dengan metode posisi IDS, IDS cenderung diurutkan sebagai kerangka kerja berbasis host dan berbasis jaringan. Dalam IDS berbasis host, IDS memberi keuntungan pada setiap host yang ingin mengikutinya. IDS dapat memutuskan apakah upaya serangan berhasil dan dapat mengatasi penipuan lokal. Dalam kerangka kerja berbasis organisasi, diamati bahwa tamu organisasi dari akses yang tidak disetujui dengan panduan yang digunakan oleh host untuk membuat hubungan yang nyaman dengan kerangka kerja yang dimiliki. Komponen ini membutuhkan biaya yang lebih sedikit untuk pengaturan, dan juga mendukung untuk memilah penipuan ke dan dari lebih dari satu host (ditunjukkan pada Gambar 12.1).



Gambar 12.1 Sistem deteksi intrusi.

Kategorisasi IDS

IDS dapat dikategorikan ke dalam banyak jenis berdasarkan platform yang digunakan untuk mendeteksi serangan dan tergantung pada catatan masukan yang terkumpul dari sumber-sumber eksklusif seperti nama sistem, log audit, aktivasi pengguna atau gadget, prosedur perangkat lunak, dan pengunjung jaringan untuk evaluasi dan serangan. IDS juga dapat diklasifikasikan terutama berdasarkan jenis serangan yang dapat dideteksi melalui setiap jenis (dijelaskan dalam Tabel 12.1).

Tabel 12.1 Perbandingan jenis IDS.

IDS	Basis	Data masukan	Serangan yang terdeteksi oleh IDS
HIDS	Host	Komposisi Sistem, aktivitas aplikasi, log Sistem, proses yang berjalan pada urutan Sistem, akses dan modifikasi berkas.	Pencatatan penekanan tombol, pencurian identitas, akses tidak sah, spamming, proses berbahaya, aktivitas botnet, penggunaan adware.
WIDS	Jaringan	Paket Lalu Lintas Organisasi, Kejadian Sebelumnya, Profil Klien.	Penipuan TCP SYN, serangan paket terbagi, Cross-Site Scripting (XSS), Cross-Site Request Forgery (CSRF)
CIDS	Server Web + Host	Penggunaan normal protokol, http, protokol bahasa pertanyaan terstruktur (square), lalu lintas tingkat utilitas, dan instruksi, catatan audit, sumber informasi program yang sedang berjalan, dan pengisian log	Serangan CANCEL DOS, serangan BYE DOS, Serangan Banjir Permintaan INVITE, Spam Media, Banjir Paket RTP
ID Berbasis Hibrida	Host+Jaringan	Sesuai dengan sistem hybrid	Sesuai dengan sistem hibrida

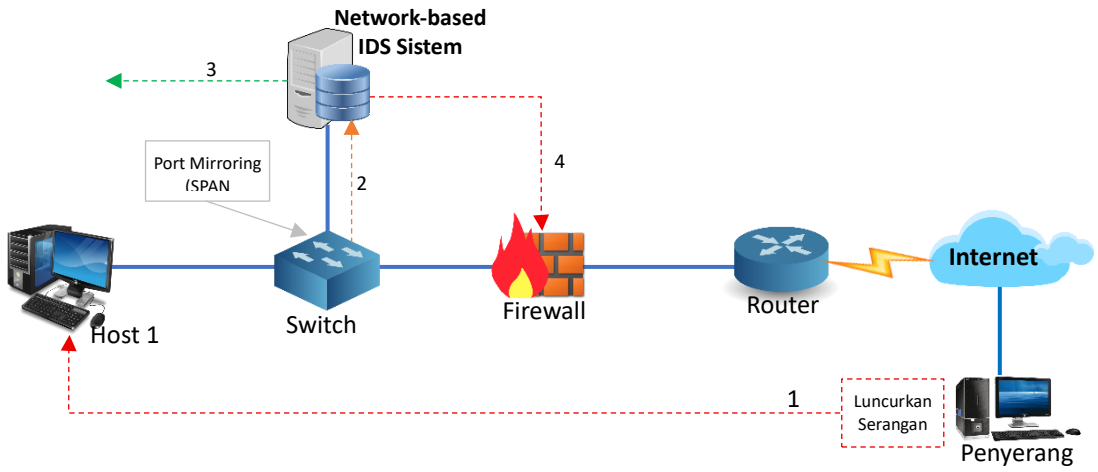
Sistem Deteksi Intrusi Web (WIDS)

Struktur deteksi intrusi web (WIDS) adalah instalasi pada faktor yang disengaja dalam jaringan untuk mengamati lalu lintas dari semua perangkat di jaringan. Sistem ini melakukan

komentar terhadap lalu lintas yang lewat di seluruh subnet dan mencocokkan pengunjung yang lewat di subnet dengan kumpulan serangan yang diketahui. Setelah serangan teridentifikasi atau perilaku aneh ditemukan, peringatan dapat dikirim ke administrator. Contoh WIDS adalah meletakkannya di subnet tempat firewall ditempatkan untuk melihat apakah seseorang mencoba membobol firewall.

WIDS biasanya digunakan atau diposisikan pada faktor strategis di seluruh komunitas, yang dimaksudkan untuk mencakup lokasi seseorang tempat pengunjung situs kemungkinan besar rentan terhadap penipuan. Umumnya, sistem ini diterapkan ke seluruh subnet, dan mencoba menyesuaikan setiap pengunjung situs yang lewat ke pustaka serangan yang diketahui. Secara pasif, ia mengamati pengunjung komunitas yang datang melalui titik-titik di komunitas tempat ia ditempatkan. Mereka mungkin sangat halus dan mungkin sulit didekati oleh penyusup. Ini menunjukkan bahwa penyusup mungkin tidak menyadari bahwa penipuan kapasitas mereka terdeteksi dengan bantuan WIDS.

Mesin deteksi intrusi berbasis jaringan menganalisis sejumlah besar pengunjung situs jaringan, yang berarti bahwa mereka terkadang memiliki spesifisitas yang rendah. Ini menunjukkan bahwa terkadang mereka mungkin melewatkan penipuan atau mungkin tidak menemukan sesuatu yang terjadi dalam lalu lintas terenkripsi. Dalam beberapa kasus, mereka mungkin memerlukan keterlibatan manual tambahan dari administrator untuk memastikan bahwa mereka dikonfigurasi dengan benar (ditunjukkan pada Gambar 12.2).



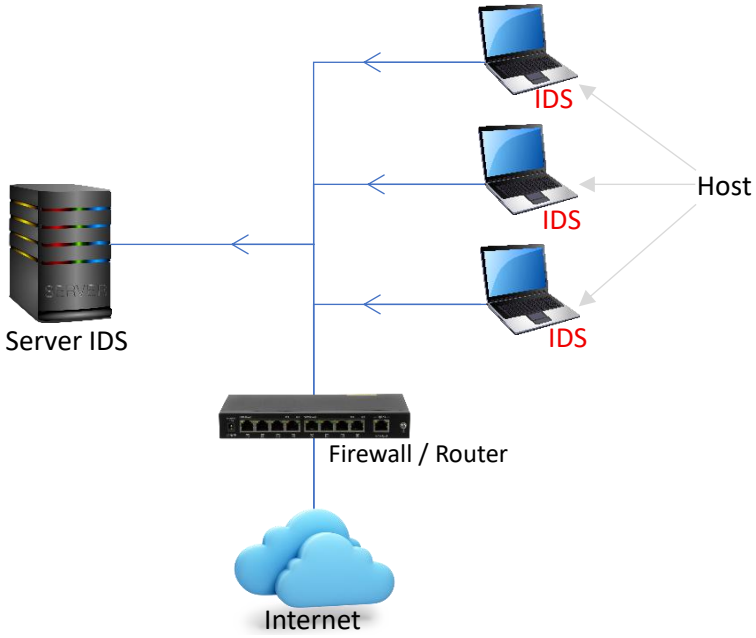
Gambar 12.2 Sistem deteksi intrusi berbasis web.

Sistem Deteksi Intrusi Host (HIDS)

Sistem deteksi intrusi host (HIDS) berjalan pada host atau gadget yang tidak memihak di jaringan. HIDS memantau paket masuk dan keluar dari perangkat saja dan akan memberi tahu administrator jika aktivitas mencurigakan atau berbahaya terdeteksi. Sistem ini mengambil foto file mesin saat ini dan membandingkannya dengan foto sebelumnya. Jika dokumen sistem analitis telah diedit atau dihapus, peringatan dikirim ke administrator untuk diteliti. Contoh penggunaan HIDS dapat dilihat pada mesin yang sangat penting bagi proyek, yang tidak diharapkan untuk mengubah tata letaknya.

HIDS berjalan pada semua gadget dalam organisasi dengan akses ke web dan bagian lain dari organisasi usaha. HIDS memiliki beberapa keunggulan dibandingkan NIDS, karena kemampuannya untuk memeriksa lalu lintas masuk dengan lebih cermat, serta berfungsi sebagai garis pertahanan kedua terhadap paket jahat yang gagal diidentifikasi oleh NIDS. Ia melihat set laporan perangkat secara lengkap dan membandingkannya dengan "snapshot" sebelumnya dari set rekaman.

Kemudian ia melihat apakah ada perbedaan besar di luar penggunaan perusahaan biasa dan memberi tahu administrator apakah ada dokumen atau pengaturan yang hilang atau diubah secara ekstensif. Ia sebagian besar menggunakan tindakan berbasis host seperti penggunaan aplikasi dan file, akses laporan di seluruh perangkat, dan log kernel. Sistem deteksi intrusi berbasis jaringan dan host adalah metode yang paling umum untuk mengekspresikan kategori ini, dan Anda juga akan menemukan NIDS yang sangat sering dikutip di ruang ini. Ia dapat dianggap hanya sebagai bentuk NIDS (ditunjukkan pada Gambar 12.3).



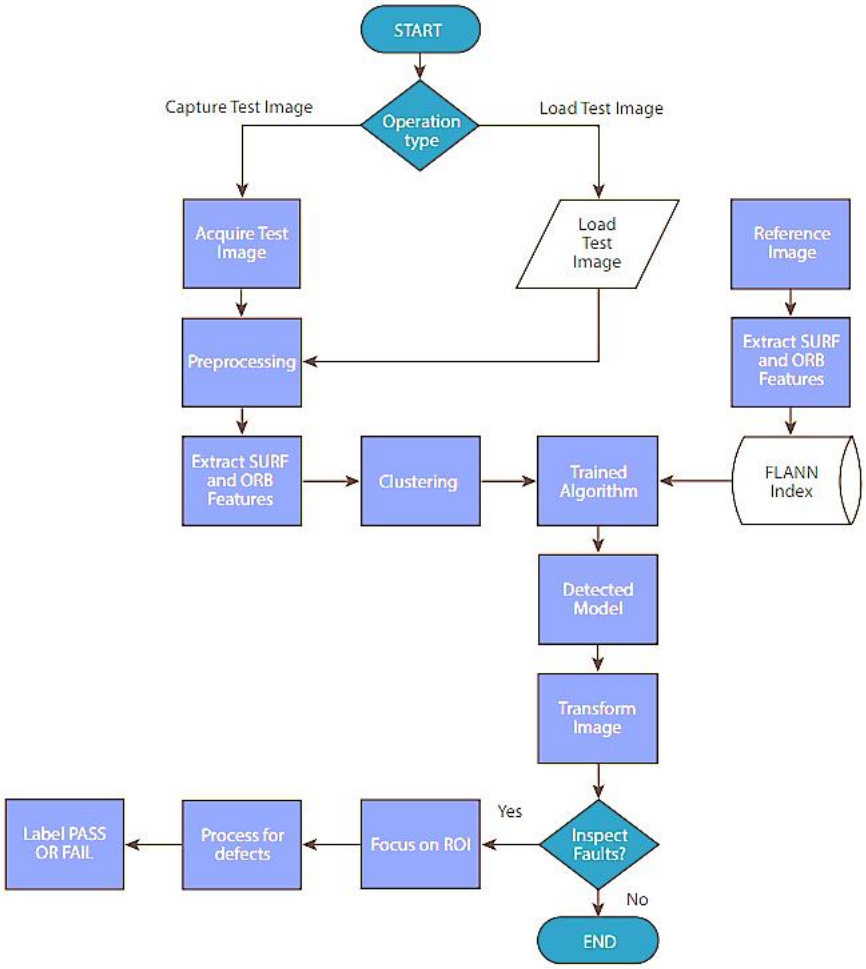
Gambar 12.3 Sistem Deteksi Intrusi Berbasis Host

Sistem Deteksi Intrusi Berbasis Kustom (CIDS)

CIDS menggabungkan perangkat atau agen yang akan selalu berada di bagian depan server, mengendalikan dan menguraikan protokol antara seseorang/alat dan server. Ia mencoba untuk memudahkan server web dengan sering melacak aliran protokol HTTPS ke dalam dan mengambil alih transportasi protokol HTTP terkait. Karena HTTPS tidak dienkripsi dan sebelum langsung masuk ke lapisan presentasi webnya, maka gadget ini mungkin perlu berada di antarmuka ini, untuk menerapkan HTTPS.

Sistem Deteksi Intrusi Berbasis Protokol Aplikasi (APIDS)

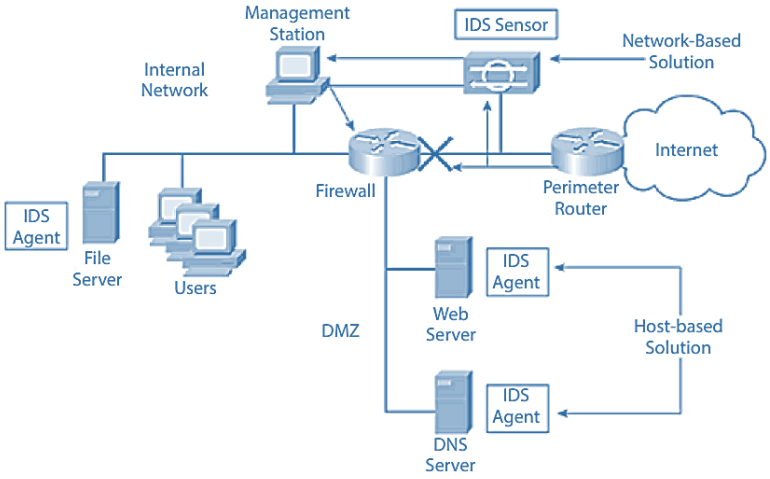
APIDS adalah sistem atau agen yang biasanya berada dalam satu set server. Ia mengidentifikasi intrusi dengan bantuan pemantauan dan penafsiran percakapan pada protokol khusus perangkat lunak. Sebagai contoh, ini dapat memantau sq. Protokol yang dinyatakan ke middleware saat bertransaksi dengan database di dalam webserver (ditunjukkan pada Gambar 12.4).



Gambar 12.4 Sistem Deteksi Intrusi Berbasis Protokol Aplikasi (APIDS).

12.2.2.5 Sistem Deteksi Intrusi Hibrida

Perangkat deteksi intrusi hibrida dibuat dengan bantuan kombinasi dua atau lebih teknik perangkat deteksi intrusi. Dalam perangkat deteksi intrusi hibrida, catatan agen host atau sistem dicampur dengan fakta komunitas untuk memperluas tampilan keseluruhan perangkat jaringan. Mesin deteksi intrusi hibrida lebih efektif dalam penilaian perangkat deteksi intrusi lainnya. Prelude adalah contoh IDS hibrida (ditunjukkan pada Gambar 12.5).



Gambar 12.5 Sistem deteksi intrusi hibrida.

12.3 PROGRAM DETEKSI

Era terkini struktur deteksi intrusi bisnis berbasis simpul elemen besar dan deteksi klien. Misalnya, perangkat masa kini sama sekali tidak memiliki kemampuan untuk mendeteksi. Kurangnya deteksi penipuan dalam struktur komersial, sebagian besar fokusnya adalah pada deteksi variasi, bukan sekadar perluasan deteksi penyalahgunaan. Sistem yang berisi dua strategi juga berguna untuk penelitian. Mengevaluasi masalah deteksi variasi diperlukan untuk menghilangkan indikasi palsu karena setiap aktivitas di luar profil yang dikenali akan memicu alarm.

Sejujurnya, harga indikasi yang salah adalah masalah terbatas dalam IDS. Kecepatan komunitas yang berlipat ganda, jaringan yang tertukar, program produk enkripsi telah menghasilkan desain yang mengarah pada penemuan lengkap berbasis perangkat. Metodologi baru yang menarik lainnya adalah penemuan gangguan terdistribusi, di mana struktur berbasis perangkat mengungkap beberapa siaran pada organisasi dan memindahkan pengetahuan yang diperiksa ke halaman situs utama.

Deteksi Penyalahgunaan

Biasanya, pembuatan ID dilakukan secara prematur dan tergesa-gesa. Di dalam domain modern, pemasok baru muncul secara rutin tetapi secara rutin diserap oleh yang lain. Di bidang pemeriksaan, penyebaran siklus sedang dieksplorasi. Namun, struktur hipotetis standar terus menjadi kurang. Metodologi penting yang telah diusulkan untuk penemuan penyalahgunaan adalah kerangka kerja induk, investigasi tanda tangan, pemeriksaan kemajuan status, dan penambangan informasi.

Pendekatan juga telah diusulkan termasuk jaring Petri yang diarsir dan pemikiran berbasis kasus. Deteksi penyalahgunaan mencari gaya penipuan yang diketahui. Ini adalah strategi yang digunakan melalui era kontemporer sistem deteksi intrusi komersial. Kerugian dari pendekatan ini adalah bahwa ia mampu mendeteksi intrusi yang paling efektif yang mematuhi pola yang telah dijelaskan sebelumnya. Proses terpenting yang diperkenalkan untuk deteksi anomali adalah ES (*Expert System*).

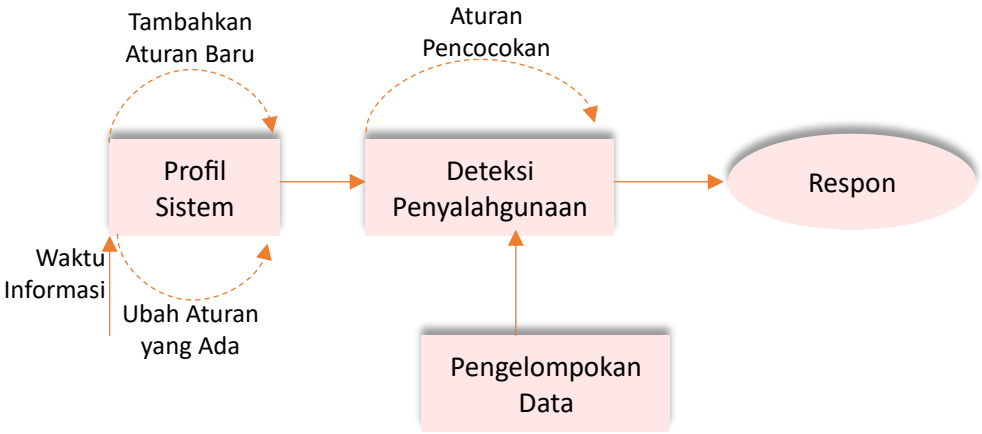
Expert System

ES menangani lokasi ilegal dengan memanfaatkan serangkaian aturan tetap untuk menggambarkan kesalahan representasi. Peristiwa tinjauan diubah menjadi catatan yang memiliki signifikansi semantik di dalam mesin pakar. Motor induksi kemudian menarik ujung-

ujungnya dengan menggunakan aturan dan data ini. Contoh ES tersebut adalah GIDS (*Grid Intrusion Detection System*), C-BEST (*Construction Based Expert System*), dll.

IDX adalah kerangka kerja utama pengenalan gangguan model berbasis informasi untuk Unix System V. Kerangka kerja ini menggabungkan informasi tentang kerangka kerja objektif, profil riwayat latihan klien sebelumnya, dan heuristik identifikasi gangguan. Hasilnya adalah kerangka kerja berbasis informasi yang dilengkapi untuk mengidentifikasi pelanggaran eksplisit yang terjadi pada kerangka kerja objektif. Komponen khusus GIDS adalah menggabungkan realitas yang menggambarkan kerangka kerja objektif dan heuristik yang dilambangkan dalam keputusan yang mengidentifikasi pelanggaran tertentu dari jejak tinjauan kerangka kerja objektif. IDX selanjutnya menjalankan kerangka kerja bawahan.

C-BEST (dibuat di SRI) adalah master forward-binding berbasis standar, kerangka kerja yang telah diterapkan untuk penemuan interupsi berbasis tanda tangan selama bertahun-tahun. Ide utamanya adalah untuk menunjukkan kualitas perilaku yang merugikan dan kemudian memantau aliran kejadian yang dihasilkan oleh gerakan kerangka kerja, yang bertujuan untuk mengenali tanda tangan interupsi. C-BEST adalah lapisan kerangka kerja master yang dapat diperoleh secara luas, yang menggunakan bahasa definisi standar yang cukup mudah digunakan oleh non-spesialis. Kerangka kerja tersebut pertama kali disampaikan dalam kerangka kerja ID MIDAS di Pusat Keamanan Komputer Nasional (NCSC). Kemudian, C-BEST dipilih sebagai motor induksi berbasis standar NIDES, pengganti model IDES. Cangkang kerangka kerja master C-BEST juga digunakan dalam master EMERALD, motor investigasi tanda non-eksklusif (ditunjukkan pada Gambar 12.6).



Gambar 12.6 Sistem pakar deteksi penyalahgunaan (MDES).

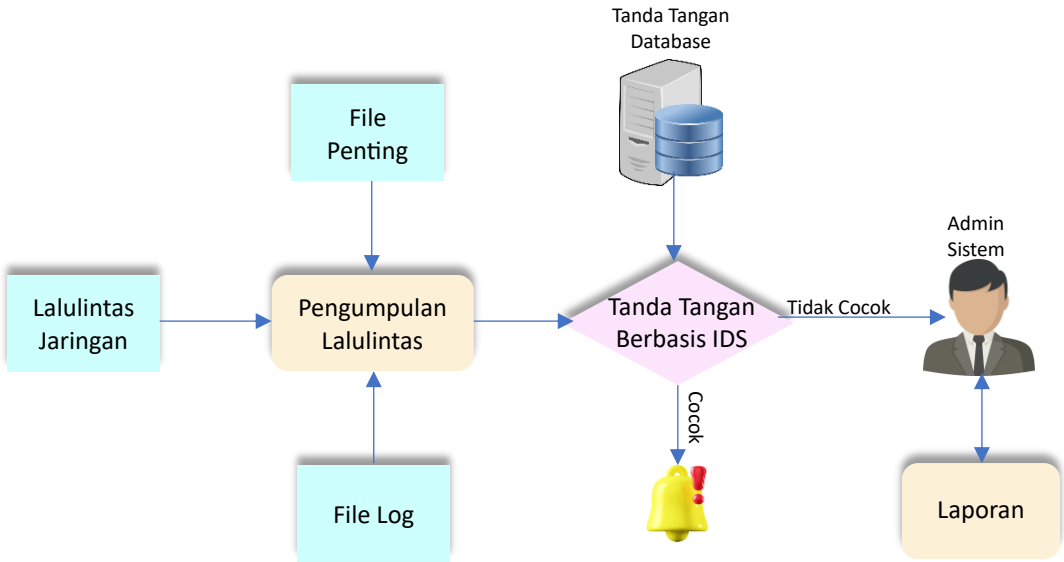
Analisis Prangko

Penilaian tanda mengubah penggambaran semantik pemerasan menjadi catatan yang mungkin ditemukan di dalam jalur tinjauan dengan cara yang wajar. Contoh wawasan tersebut mencakup rangkaian peristiwa tinjauan yang dibuat oleh pemerasan atau contoh informasi yang dapat dicari di dalam jalur tinjauan. Kerangka kerja yang menggunakan investigasi tanda tangan mencakup Haystack, NetRanger RealSecure, dan MiSig (*Misuse Signatures*).

- Stack adalah perangkat deteksi penyalahgunaan yang memungkinkan petugas keamanan angkatan udara untuk menemukan penyalahgunaan mainframe Unisys. Bekerja pada catatan jalur audit yang dikurangi, ia melakukan deteksi penyalahgunaan

sepenuhnya berdasarkan batasan perilaku yang diberlakukan melalui peraturan keselamatan yang tepat dan pada model perilaku konsumen tradisional.

- Net-Ranger terdiri dari dua modul: sensor dan eksekutif. Sensor adalah unit tayangan video keamanan jaringan yang memeriksa lalu lintas organisasi pada segmen organisasi dan pengukuran pencatatan yang dibuat dengan panduan switch cisco untuk mengejutkan serangan berbasis organisasi. Kepala bertanggung jawab atas kendali sekelompok sensor dan dapat didasarkan secara progresif untuk mengendalikan organisasi besar seperti yang ditunjukkan pada Gambar 12.7.
- Keamanan nyata (maju pada struktur keamanan web) terdiri dari tiga modul: motor jaringan, pengiklan mesin, dan kepala. Motor organisasi adalah unit tayangan video jaringan yang dilengkapi dengan tanda-tanda misrepresentasi yang dikoordinasikan terhadap tamu pada hyperlink organisasi. Dealer gadget memiliki kerangka kerja pengenalan gangguan berbasis yang menunjukkan catatan log keamanan yang rumit pada sejumlah. Modul-modul tersebut melaporkan temuan mereka kepada administrator penting, yang menunjukkan catatan kepada individu dan menyajikan fungsionalitas untuk dealer gadget administrasi jarak jauh dan motor jaringan.
- Missing menerapkan bahasa tingkat tinggi untuk tanda tangan abstrak. Ia mencoba untuk mengatasi keterbatasan tertentu dari sistem deteksi penyalahgunaan tradisional, yang mencakup ekspresi terbatas tanda tangan yang diungkapkan dalam bahasa tingkat rendah dan algoritma pelacakan tetap untuk penyalahgunaan yang mengalami kesulitan beradaptasi dengan lingkungan berjalan yang berubah atau tujuan keamanan. Melalui bahasa tingkat tingginya, missing dapat membentuk penyalahgunaan dalam bentuk yang mudah dengan ekspresi tinggi.



Gambar 12.7 Analisis berbasis tanda tangan dalam IDS.

Penambangan Data

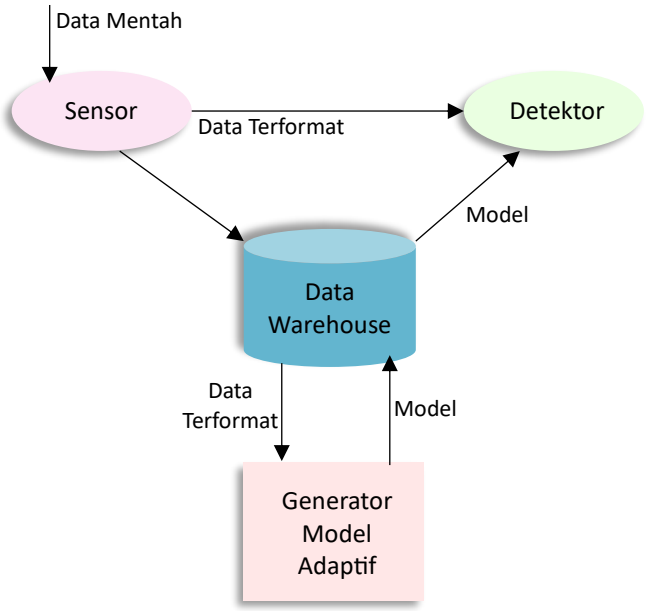
Pengurutan statistik (informasi) lebih menyukai cara penggalian yang tidak sepele dari informasi internal, yang sebelumnya tidak disebutkan namanya, dan yang mungkin berguna dari basis data. Model untuk deteksi intrusi), dan penemuan otomatis petunjuk prediktif

ringkas untuk deteksi intrusi. IDS menggunakan pengurutan (penambangan) yang mencakup JAM. MARAD dan petunjuk pencarian otomatis untuk sistem IDS. Sistem deteksi anomali instans menggunakan pengurutan informasi yang mencakup JAM dan MARAD.

JAM menggunakan teknik penambangan fakta untuk menemukan pola intrusi. Kemudian menerapkan pengklasifikasi meta-bacaan untuk memeriksa tanda-tanda serangan. Kumpulan aturan asosiasi menentukan hubungan di antara bidang-bidang dalam data jalur audit dan episode umum dari serangkaian pedoman membentuk gaya berurutan dari aktivitas audit. Fitur kemudian diekstraksi dari setiap algoritma dan digunakan untuk menghitung model perilaku intrusi. Pengklasifikasi menyusun tanda-tanda serangan.

Jadi pada dasarnya, penambangan catatan dalam JAM membangun model deteksi penyalahgunaan. JAM menggunakan strategi penambangan fakta untuk menemukan gaya intrusi. Kemudian menerapkan pengklasifikasi meta-bacaan untuk menganalisis tanda-tanda serangan. Algoritme aturan afiliasi menentukan hubungan antara bidang di dalam statistik jejak audit dan episode umum dari serangkaian model aturan yang berurutan dari berbagai jenis aktivitas audit. Fitur kemudian diekstraksi dari setiap algoritme dan digunakan untuk menghitung model perilaku intrusi. Pengklasifikasi menyusun tanda-tanda penipuan. Jadi pada dasarnya, penambangan data di JAM membangun model deteksi penyalahgunaan.

JAM membuat pengklasifikasi menggunakan protokol yang mempelajari aplikasi pada statistik pelatihan penggunaan mesin. Perangkat diperiksa dengan informasi dari penipuan berbasis email, dan dengan serangan jaringan menggunakan statistik dump TCP. Identifikasi MARAD menggunakan penambangan fakta untuk membuat kebijakan untuk deteksi penyalahgunaan. Alasannya adalah bahwa struktur saat ini memerlukan upaya manual yang besar untuk memperluas peraturan untuk deteksi penyalahgunaan.



Gambar 12.8 Penambangan data dalam sistem deteksi intrusi.

MARAD ID menerapkan penambangan informasi untuk meninjau informasi guna membuat model yang secara tepat menangkap standar perilaku interupsi dan latihan khas (ditunjukkan pada Gambar 12.8).

12.4 POHON KEPUTUSAN

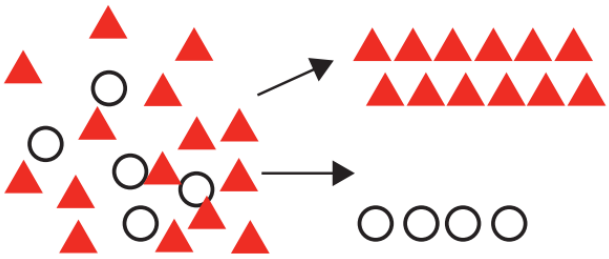
Penambangan statistik (informasi) adalah teknik untuk menemukan informasi dari kumpulan data besar yang didukung oleh informasi dan teknik kecerdasan buatan untuk memecahkan masalah kehidupan nyata yang rumit. Pohon pilihan adalah jenis metode penting dalam kelas penambangan fakta. Pohon preferensi digambarkan sebagai bentuk seperti diagram alir atau seperti pohon dari berbagai fakta khusus. Pohon keragu-raguan, setiap simpul internal mewakili pemeriksaan pada suatu fitur, sementara setiap batang mewakili konsekuensi terakhir dari pengujian dan setiap simpul daun mewakili label kelas.

Rute dari simpul dasar ke simpul daun mewakili bentuk peraturan. Dari sudut pandang deteksi intrusi, algoritme kelas dapat membedakan statistik komunitas sebagai penipuan, jinak, pemindaian, atau beberapa kategori minat lainnya menggunakan catatan seperti port pengiriman/tujuan, alamat IP, dan rentang byte yang dikirim dalam beberapa waktu yang tidak ditentukan di masa mendatang dari suatu koneksi. Pengklasifikasi pohon ekspansi memiliki bentuk yang mudah yang menyimpan dan memberi label informasi baru dengan sempurna. Pengklasifikasi dalam pohon keputusan berisi banyak algoritme, seperti CART, C4.5, dan ID3.

Pohon Klasifikasi dan Regresi (CART)

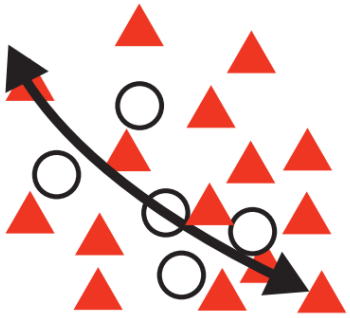
Pohon Keputusan biasanya digunakan dalam penambangan data dengan tujuan membuat model yang memprediksi biaya target (atau variabel yang ditetapkan) berdasarkan nilai beberapa masukan (atau variabel yang tidak bias). Dalam postingan terkini, kita akan membahas metodologi pohon keputusan CART. Teknik CART atau pohon klasifikasi dan regresi diperkenalkan pada tahun 1984 oleh Leo Breiman, Jerome Friedman, Richard Olsen, dan Charles Stone.

CART membangun pohon biner yang berarti kumpulan data yang berisi simpul paling efektif dapat dibagi dalam organisasi. CART dapat menangani beberapa bentuk data seperti setiap informasi unik dan aritmatika. CART menggunakan indeks Gini untuk memutuskan atribut. Fungsi dengan diskon paling penting dalam ketidakmurnian digunakan untuk membagi simpul kumpulan data. Fungsi ini menggunakan pemangkasan kompleksitas biaya dan juga menghasilkan pohon regresi. Pohon Klasifikasi: di mana variabel sasaran bersifat spesifik dan pohon tersebut digunakan untuk mengidentifikasi "Kelas" di mana variabel sasaran kemungkinan besar dapat masuk (Gambar 12.9).



Gambar 12.9 Pohon klasifikasi.

Waktu regresi: Di mana variabel sasaran tidak berhenti dan pohon tersebut digunakan untuk mengharapkan nilainya (Gambar 12.10).



Gambar 12.10 Pohon regresi.

Dikotomi Iteratif 3 (ID3)

Rangkaian pedoman ini sebagian besar didasarkan pada standar pisau cukur Occam, dengan konsep menciptakan pohon preferensi terkecil dan paling hijau. ID3 menggunakan perolehan statistik dari setiap karakteristik untuk pohon pilihan pengembangan. Bakat yang memiliki manfaat yang sangat bagus dapat memilih data statistik terdistribusi. Rangkaian aturan ID3 memiliki beberapa kelemahan, yang meliputi untuk beberapa waktu, statistik dapat dikategorikan secara berlebihan, dan hanya satu komponen pada satu waktu yang dipertimbangkan untuk membuat pohon keputusan. Karakteristik yang paling efektif pada suatu waktu diperiksa sehingga akan membuat pilihan, dan tidak lagi berurusan dengan karakteristik tanpa henti beserta biaya tersembunyi untuk membuat pohon.

Rangkaian kebijakan ini berubah menjadi berbasis desain sepenuhnya pada persyaratan pisau cukur Occam, pemikiran untuk membuat pohon keinginan hijau minimum dan maksimum. ID3 menggunakan perolehan informasi dari setiap fungsi untuk pengembangan pohon pilihan. Bakat yang memiliki keuntungan yang sangat baik dapat memilih untuk membagi informasi dan catatan. Rangkaian pedoman ID3 memiliki kelemahan, yang meliputi untuk sementara waktu, statistik dapat diberi label berlebihan, hanya satu karakteristik pada suatu waktu yang dipertimbangkan untuk pohon keputusan. Hanya satu karakteristik pada suatu waktu yang diperiksa dengan tujuan untuk membuat pilihan, dan tidak lagi berurusan dengan karakteristik tanpa henti serta biaya yang hilang untuk membuat pohon.

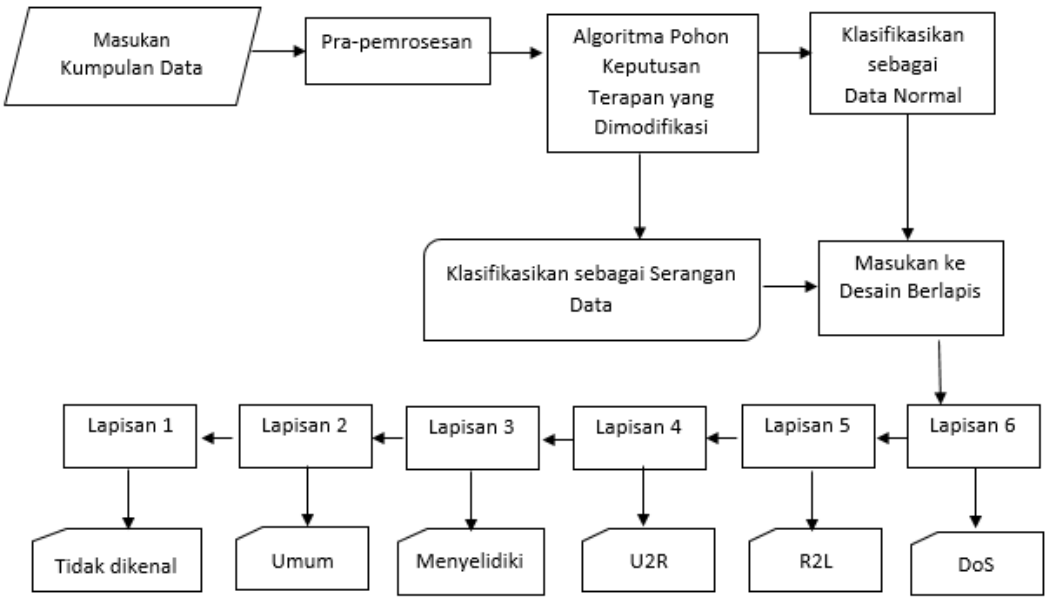
C 4.5

C4.5 adalah perluasan dari rangkaian aturan ID3 dan pengklasifikasi aritmatika. Ia mengatasi kesulitan yang terkait dengan kalkulasi ID3 seperti beradaptasi dengan rekaman non-inferensi dan kualitas yang hilang. Ia mengikuti gawai yang sama seperti ID3 untuk rekaman tertentu dan menggunakan teknik rasio faktor untuk jenis rekaman numerik.

12.5 MODEL PENAMBANGAN DATA UNTUK MENDETEKSI SERANGAN

Dalam metode pendeteksian serangan ini digunakan kumpulan data KDD99. Kumpulan data ini berisi 42 fitur termasuk kolom kelas yang berisi data jenis umum dan serangan. Model ini melalui dua fase berbeda. Di dalam bagian pertama, fakta jaringan diproses terlebih dahulu untuk mengubah bentuk informasi yang terpisah menjadi rekaman jenis integer dan setelah konversi; pohon pilihan dibangun dengan bantuan kumpulan fakta yang diproses terlebih dahulu. Kumpulan informasi yang diproses terlebih dahulu ini cukup mampu membedakan informasi rekaman yang dikenal dan rekaman serangan di dalam simpul balita pohon.Fase lain

dikenal sebagai "Fase Deteksi". Bagian ini menganalisis serangan yang terkait dengan kelasnya dan berbagai prevalensinya sendiri. Bagian ini mengidentifikasi rekaman yang berisik atau fakta yang hilang yang disebut serangan (dijelaskan dalam Gambar 12.11).



Gambar 12.11 Model algoritma pohon keputusan yang dimodifikasi.

Kerangka Teknik

Model yang diperkenalkan adalah model yang disempurnakan dari kumpulan aturan pohon C4.5 yang menangani informasi non-linier dan spesifik secara bersamaan untuk kumpulan data terdistribusi seperti setiap hari dan penipuan pada tingkat daun. Dalam C4.5 sederhana, kumpulan data yang merupakan kumpulan data memerlukan tata letak yang lebih pendek, dan mengelola catatan unik dan terus-menerus satu demi satu yang merupakan metode yang memakan waktu, dan juga memutuskan nilai pemecahan merupakan kesulitan penting dalam membuat pohon keputusan.

Dengan mengkhususkan diri dalam situasi seperti itu, kami telah mengubah beberapa contoh dalam model yang kami perkenalkan, seperti sebagai pengganti mengelola informasi eksplisit dan non-prevent satu demi satu, kami mengubah catatan eksplisit menjadi fakta berkelanjutan dalam pra-pemrosesan, dan tanpa menopang kumpulan data, kami segera mempraktikkan kumpulan aturan untuk klasifikasi. Untuk kemampuan membagi. Langkah-langkah algoritmanya adalah sebagai berikut:

Algoritma

- Input: Kumpulan Data Acak (D)
- Jumlah total sampel dalam D: N_R
- Jumlah elemen unik: N_U
- Kolom dalam kumpulan data: C_D
- Nilai berbeda yang ada dalam C_D: V
- Elemen dalam C_D: D_i
- Jumlah total elemen unik dalam N_U: T_i
- Output: Data Terklasifikasi

Awal

Langkah 1.

Jika kumpulan data input memiliki tipe kelas yang sama, maka Leaf ← Class_Name

Langkah 2.

Jika kelas tertentu ada dalam data input

$$\text{Leaf} \leftarrow \text{Histogram}$$

Langkah 3. Entropi Dataset

$$En(N_R) = \sum_{i=1}^{UC} \text{freq} \frac{(Ti, NR)}{|NR|} * \frac{\log freq(Ti, NR)}{|NR|}$$

Langkah 4. Informasi setiap atribut

$$(\text{Info}_{\text{Att}}(\text{N}_\text{R}) = \sum_{i=1}^{V} |\frac{Di}{NR}| * En(Di)$$

Langkah 5. Informasi Gain

$$(\text{IG}\ (\text{N}_\text{R})) \ = \ \text{En}\ (\text{N}_\text{R}) - \text{Info}_{\text{att}}(\text{N}_\text{R})$$

Langkah 6. Pembagian Informasi

$$(\text{Div\_Info}(\text{N}_\text{R})) = \sum_{\text{i=1}}^{\text{V}} \frac{|\text{Di}|}{\text{NR}} * \log_{\text{W}} \frac{|\text{Di}|}{\text{NR}}$$

Langkah 7. Rasio Gain

$$(\text{G}_{\text{Ratto}}(\text{N}_\text{R}))= \frac{\text{IG}}{\text{DIv Info (NR)}}$$

Langkah 8 : Node Putusan ← Atribut rasio gain absolut

Langkah 9 : Membagi nilai ← Rata-rata

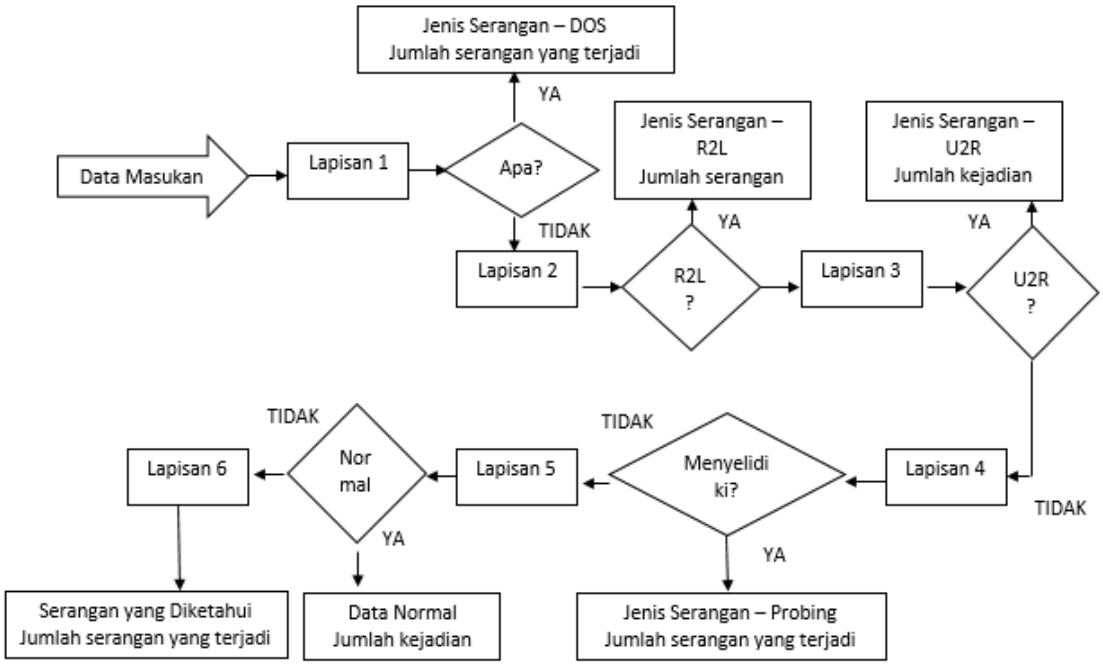
Subset Kiri ← (Dataset< Nilai yang dibagi)

Subset Kanan ← (Dataset>Nilai yang dibagi)

Langkah 10 : Ulangi langkah 1 hingga 9 pada setiap subset yang dihasilkan dengan bantuan pembagian set menjadi atribut dan masukkan node yang satu sebagai turunan dari node induk.

Selesai

Setelah berhasil menghasilkan pohon seleksi, kita perlu mengklasifikasikan regular kecurangan dengan kelasnya, yang terdiri dari dos atau U2R, R2l, atau jenis probe (ditunjukkan di bawah pada Gambar 12.12).



Gambar 12.12 Deteksi aliran serangan.

Uji set data masukan mengalami enam tingkatan di mana pada setiap konfirmasi contoh-contoh tersebut jika dibandingkan dalam ketahanan dengan jenis gaya penipuan. Setiap kemegahan penipuan berisi nama serangan yang terkait dengan kemegahan tersebut. Hasil uji setiap lapisan menampilkan berbagai pola yang terkait dengan kemegahannya. Mari kita asumsikan bahwa pola diubah menjadi struktur lapisan, lalu tingkatkan jumlah prevalensi jenis R2L jika tidak maka uji apakah itu U2R jenis tersebut dan kemudian jenis penyelidikan.

Jika sampel tidak lagi termasuk dalam keempat jenis serangan ini maka uji apakah itu catatan jenis biasa atau tidak. Jika ya maka tingkatkan rentang kemunculan jenis biasa, dan jika tidak lagi termasuk dalam salah satu kelas tersebut, maka jadikan sebagai jenis serangan yang tidak diketahui dan perhatikan rentang kemunculannya. Pada akhirnya, hasil akhir menunjukkan bentuk sampel yang luas yang ditemukan di setiap jenis elegan. Di setiap lapisan, fakta-fakta difilter ke jenis kelas yang sesuai. Kategori luncuran serangan ditunjukkan pada jenis serangan induk 11 dengan jumlah kejadiannya.

Keputusan adalah salah satu strategi yang hebat dan terkenal untuk mendeteksi struktur. Ini membuat pilihan yang tepat tentang apa yang ada di komunitas atau tidak di situs internet komunitas, fakta pengunjung situs web adalah serangan dan catatan biasa. Model yang diperkenalkan membuat pohon keputusan dengan bantuan rasio manfaat dan rata-rata geometris untuk membagi kumpulan data. Ini juga secara efisien mengidentifikasi jenis-jenis item penipuan tertentu dalam kumpulan data dengan identifikasi statistik yang tidak diketahui.

Hasil dari versi yang diusulkan dibandingkan dengan strategi DT yang berbeda seperti ekstensi keranjang dan ID3 (C4.5) dengan bantuan kumpulan data KDD cup-99 dan versi yang diusulkan memberikan akurasi 99% untuk identifikasi serangan dengan waktu yang lebih sedikit. Keuntungan dari model yang diusulkan dibandingkan dengan C4.5 adalah risiko untuk mendapatkan biaya deteksi yang tidak wajar atas berbagai jenis penipuan dengan tingkat kesalahan dan waktu yang jauh lebih sedikit. Lukisan takdirnya adalah untuk memeriksa

kinerja keseluruhan versi ini melalui kumpulan data besar dan juga untuk mengatasi jenis penipuan yang tidak diketahui dalam sistem kontrol otomatis.

BAB 13
OPTIMASI PANEN & MONITORING JAGUNG BERBASIS IOT FIREFLY

Tantangan utama yang dihadapi oleh organisasi terkait pertanian adalah mengidentifikasi tanaman optimal yang akan menghasilkan laba lebih tinggi berdasarkan kondisi iklim yang berubah secara dinamis. Prediksi tanaman optimal mencakup produksi tanaman, pemasaran, rantai pasokan, penyimpanan, transportasi, dll., beserta kendala dan pemenuhan risiko yang terkait dengannya. Dalam penelitian terkini ini, algoritma firefly telah digunakan untuk mengoptimalkan hasil panen jagung dengan mempertimbangkan berbagai kendala dan risiko. Penelitian ini menyelidiki pengembangan modul algoritma kunang-kunang baru untuk memprediksi kondisi iklim yang optimal dan memprediksi hasil budidaya tanaman. Sebagai pra-pemrosesan, data budidaya tanaman jagung selama 96 bulan telah dikumpulkan dan diberikan sebagai respons terhadap perangkat lunak Minitab untuk merumuskan persamaan relasional.

Data yang dikumpulkan telah disimpan di cloud menggunakan IoT dan cloud harus diperbarui secara berkala untuk mendapatkan hasil yang akurat dari algoritma. Persamaan ini telah menggunakan fungsi kebugaran untuk memberikan perkiraan hasil panen yang tepat. Variabel yang perlu diperhatikan adalah jumlah rata-rata curah hujan, fasilitas irigasi, dan suhu udara atmosfer untuk mengidentifikasi kombinasi terbaik yang dapat menghasilkan budidaya tanaman jagung yang lebih tinggi. Kinerja modul yang dikembangkan telah ditemukan memuaskan dan dapat digunakan untuk prediksi tanaman lainnya juga.

13.1 PENDAHULUAN

Peramalan laba hasil panen telah menjadi masalah yang membingungkan bagi organisasi terkait pertanian dan petani untuk dipecahkan, karena hasil panen bergantung pada kolaborasi multifaset antara tanah, iklim, udara, curah hujan, irigasi, air, kelembaban, dan jenis tanaman yang dibudidayakan di dalamnya, yaitu, kombinasi yang tepat dari parameter bio-ekosistem yang diperlukan untuk hasil panen terbaik. Di sisi lain, parameter ini bervariasi secara dinamis dan sulit diprediksi sebelumnya dan setiap penyimpangan dalam prediksi menghasilkan laba yang lebih rendah.

Jadi, muncul kebutuhan akan model yang komprehensif untuk memprediksi dan mengidentifikasi tanaman optimal yang akan menghasilkan budidaya yang lebih baik dan pada gilirannya menghasilkan laba, yang dapat dimungkinkan melalui model matematika dan rekayasa. Begitu banyak penelitian telah dilakukan oleh para peneliti terkait pangan dan pertanian dalam peramalan tanaman dan dalam seni memprediksi hasil panen. Juga secara umum diterima oleh para peneliti bahwa prediksi dan identifikasi optimal produksi tanaman sebelum panen harus dilakukan terlebih dahulu.

Sebaliknya, identifikasi optimal tanaman dan filosofi peramalan merupakan tugas yang sulit karena melibatkan beberapa kategori pengumpulan data dari berbagai sumber seperti data meteorologi, data agronomi, data terkait kekuatan tanah, data ketersediaan air, data jenis tanaman, statistik pertanian, data iklim, data kelembapan, data intensitas cahaya matahari, curah hujan, dan data pasokan air. Meskipun beberapa indeks untuk variabel-variabel ini telah

diturunkan dalam menentukan hasil panen, kinerja respons belum berada pada tingkat yang memuaskan. Jadi, persyaratan model untuk mengidentifikasi tanaman optimal dan kondisi yang akan menghasilkan keuntungan maksimum telah muncul.

Identifikasi hasil panen optimal yang tepat waktu dan akurat sangat penting untuk budidaya, penyimpanan, transportasi, dan pemasaran tanaman maksimum yang akan menghasilkan keuntungan maksimum. Jadi, menjadi penting untuk membuat model memahami perilaku stokastik hasil panen dari data sebelumnya di semua tingkatan. Jadi dalam penelitian ini, data selama 96 bulan terakhir telah dikumpulkan dari para petani di desa Karur di distrik Namakkal, Tamilnadu, India. Yang juga diperlukan adalah evaluasi dan identifikasi tepat waktu terhadap tanaman potensial yang akan menghasilkan keuntungan maksimal, dan pada gilirannya dapat berdampak ekonomi pada produk pertanian di pasar. Jadi model yang dikembangkan dapat memberikan prediksi tanaman optimal yang akan menghasilkan keuntungan maksimal setiap bulan dan setiap tahun.

Para peneliti juga tertarik mengembangkan model yang dapat bekerja di seluruh dunia berdasarkan data geolokasi dan data agronomi, yang dapat memperkirakan hasil panen terbaik menggunakan pendekatan seperti regresi linier berganda, jaringan syaraf tiruan, pencarian Tabu, dll. Jadi tujuan utama penelitian ini adalah menggunakan bank data awan melalui *Internet of Things* (IoT) untuk menyimpan catatan sebelumnya dan menyimpan data terkini untuk identifikasi dan prediksi panen optimal yang akan menghasilkan keuntungan maksimal.

Pengguna akhir utama model yang dikembangkan adalah petani lokal yang terlibat langsung dalam budidaya tanaman. Keputusan akan diambil sesuai dengan pengalaman, hasil model yang dikembangkan, dan rekomendasi yang diberikan oleh Kementerian Pertanian. Di tingkat pertanian, model yang dikembangkan harus memiliki kemampuan untuk mengidentifikasi kekurangan yang dapat dikendalikan dan dengan kepuasan, akan menghasilkan keuntungan maksimal. Jadi model yang dikembangkan harus tangguh dan mampu mengakomodasi semua kendala dan tujuan secara bersamaan untuk memberikan solusi yang layak. Oleh karena itu dalam penelitian ini, algoritma kunang-kunang telah digunakan untuk mengenali tanaman ideal yang dapat menghasilkan pendapatan ekstrim.

13.2 SURVEI LITERATUR

Sebelumnya, simulasi dan regresi merupakan dua model utama untuk memprediksi panen optimal. Model simulasi dicirikan melalui hubungan ilmiah dan keakuratan hasil bergantung pada data yang tersedia, tetapi data pertanian mungkin jarang dan tidak lengkap sehingga model simulasi mungkin tidak sesuai untuk aplikasi tersebut. Di sisi lain, model regresi telah digunakan dalam skala besar. Beberapa peneliti menggunakan beberapa model regresi untuk memprediksi prediksi hasil panen agro meteorologi. Selain itu, banyak model heuristik baru yang digunakan untuk memprediksi hasil panen oleh para peneliti. Beberapa peneliti secara khusus mengidentifikasi parameter yang memengaruhi hasil panen dan mengoptimalkan parameter tersebut seperti sifat tanah.

Meskipun banyak penelitian telah dilakukan, sebagian besar belum mengakomodasi masalah multi-kendala dan multi-objektif dengan solusi alternatif dengan mengendalikan variabel. Jadi satu-satunya solusi adalah penggunaan algoritma evolusioner. Di antara berbagai algoritma mimik, algoritma kunang-kunang telah diidentifikasi sebagai algoritma terbaik yang

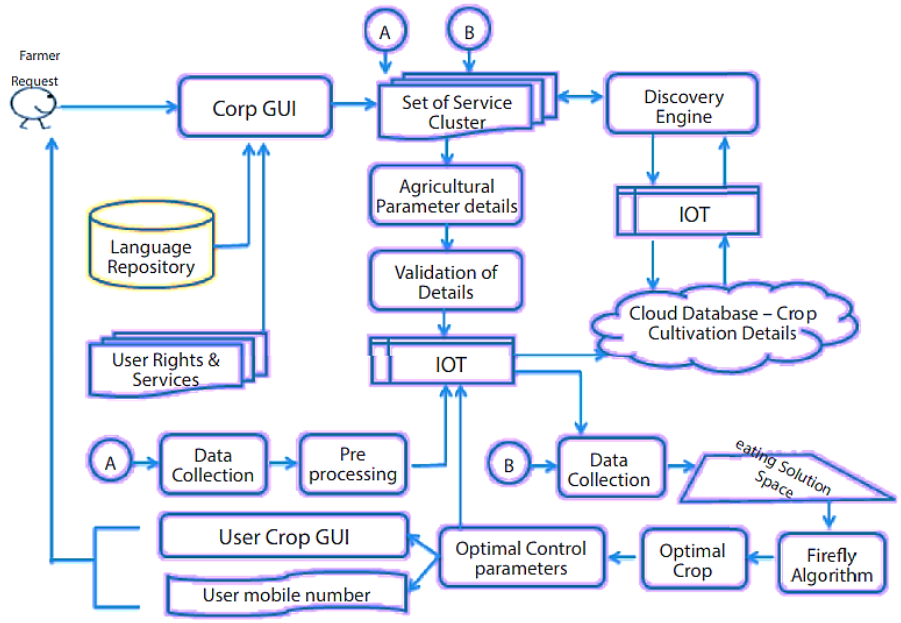
akan menghasilkan hasil yang lebih baik. Selain kondisi cuaca, budidaya tanaman juga bergantung pada parameter seperti penyakit, serangan hama, perencanaan operasi panen, dll. Jadi model optimasi memerlukan manajemen yang efektif dari faktor-faktor ini juga.

Beberapa peneliti mempertimbangkan benih, pupuk, agrokimia dan mesin pertanian untuk estimasi produksi tanaman. Beberapa peneliti mempertimbangkan kondisi cuaca yang tidak menentu dan tidak dapat diprediksi seperti badai yang memengaruhi hasil pertumbuhan tanaman. Model yang umum diterapkan adalah model regresi linier, model regresi nonlinier dan model regresi polinomial. Para peneliti juga mempertimbangkan perubahan iklim untuk mensimulasikan hasil kinerja jagung. Persamaan regresi linier berganda juga telah dibangun untuk mengkorelasikan variabel hasil dan respons. Banyak peneliti yang membahas data curah hujan, suhu permukaan, kelembapan, dan indeks vegetasi dari berbagai negara bagian dan mengumpulkan data selama bertahun-tahun.

Penelti melakukan eksperimen dengan menganalisis enam lokasi berbeda yang memiliki perubahan iklim yang beragam di Eropa dan Selandia Baru yang berfokus pada penerapan nitrogen dan strategi pengendalian gulma menggunakan ANN dan Pohon Keputusan. Namun, model pohon keputusan memiliki keterbatasan seperti hasil yang tidak stabil, pertimbangan variabel yang lebih sedikit, dll. Jadi dalam penelitian ini algoritma kunang-kunang telah digunakan untuk mengoptimalkan dan mengidentifikasi tanaman terbaik yang akan menghasilkan keuntungan yang lebih baik dan budidaya yang maksimal.

13.3 KERANGKA EKSPERIMEN

Setelah meninjau metodologi yang ada terkait dengan hasil panen, bagian ini membahas penerapan algoritma kunang-kunang untuk mengidentifikasi tanaman optimal untuk menghasilkan keuntungan maksimal. Seluruh pekerjaan penelitian telah dibagi menjadi beberapa sub-kategori seperti Pengumpulan data, Praproses data, Perumusan IoT dan basis data cloud, Pertimbangan dan asumsi awal, Perumusan persamaan relasional, Implementasi algoritma firefly, Validasi dan Pembaruan cloud. Arsitektur kerangka kerja eksperimental ditunjukkan pada Gambar 13.1.



Gambar 13.1 Arsitektur kerangka eksperimental.

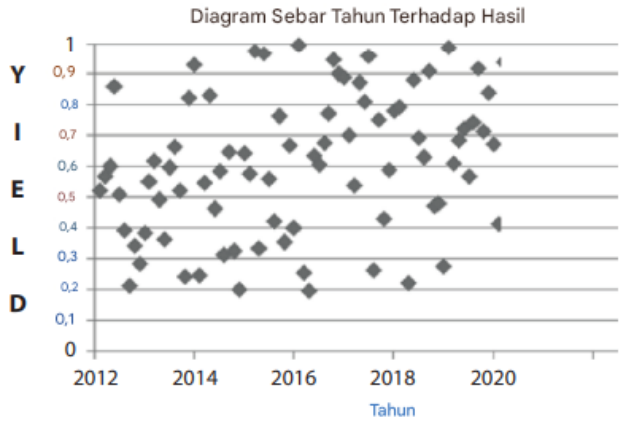
Tahap pertama adalah tahap pengumpulan data dan studi eksperimental ini telah dilakukan di desa-desa dekat Karur, Distrik Namakkal, Tamilnadu, India. Data terkait hasil panen jagung untuk musim tanam berturut-turut selama 96 bulan telah dikumpulkan dari para petani dan ahli pertanian. Data telah dikumpulkan dari survei satu lawan satu, dengan mengedarkan formulir, komunikasi lisan melalui mode tanya jawab, curah pendapat dengan para penyusun, pertemuan satu lawan satu, pengumpulan lembar data, formulir Google, dll. Data yang dikumpulkan adalah hasil panen per hektar. Setelah mengumpulkan data, data yang dikumpulkan harus divalidasi untuk kebenarannya, sehingga model yang dikembangkan dapat menghasilkan kinerja yang diinginkan. Diagram sebar untuk data yang dikumpulkan telah diformulasikan dan diplot. Diagram sebar yang diplot diberikan dalam Gambar 13.2.

Tahap kedua adalah tahap praproses data. Secara umum, data dunia nyata yang dikumpulkan melalui survei sering kali tidak lengkap, berlebihan, tidak konsisten, dan kurang dalam aspek-aspek tertentu. Jadi, muncul kebutuhan untuk menyaring data untuk kelengkapannya. Merupakan aturan praktis bahwa data yang lengkap akan memberikan hasil yang akurat dan data yang tidak tepat menghasilkan solusi yang mendekati optimal dan menghabiskan waktu tambahan.

Langkah-langkah yang diikuti dalam penelitian ini untuk memfilter data yang telah dikoreksi adalah sebagai berikut:

a) Rangkaian data dengan nilai yang hilang dihilangkan dari kumpulan data,
b) Data dengan nilai puncak ekstrem yang tiba-tiba dibulatkan ke nilai rata-rata terdekat,
c) Mengisi nilai yang hilang berdasarkan sembilan nilai data yang dikelilingi,
d) Menghaluskan data yang tidak stabil dan
e) Mengidentifikasi dan menghilangkan data yang tidak konsisten.

Tahap ketiga adalah penyimpanan basis data IoT dan Cloud. Karena berbagai keuntungan seperti model yang dikembangkan harus mengidentifikasi nilai optimal dari berbagai segmen, data harus tersedia di ruang umum, data hasil panen telah diperbarui secara berkala untuk prediksi yang akurat di masa mendatang, data dapat ditafsirkan oleh ahli pertanian untuk prediksi dan statistik lainnya. Jadi satu-satunya kemungkinan adalah menambahkan data dalam penyimpanan cloud dan mengambilnya kembali untuk pemrosesan lebih lanjut.



Gambar 13.2 Diagram sebar data yang dikumpulkan.

Tahap keempat adalah pertimbangan awal untuk aplikasi. Untuk memperkecil kerumitan komputasi, beberapa parameter dengan data yang tidak lengkap dan kategori yang dapat

dikontrol diasumsikan sebagai berikut, benih berkualitas baik, tidak ada badai, pupuk yang cukup, komunikasi dan teknologi yang tepat, suhu yang seragam, tanah lempung berpasir, pengendalian gulma dan hama, pembajakan dan penggaruan. Tahap kelima adalah perumusan persamaan empiris yang menggambarkan hubungan antara parameter input dan output. Jadi, mengoptimalkan unit yang berbeda secara bersamaan membutuhkan lebih banyak kerumitan dan dengan demikian parameter dioptimalkan dan diberikan sebagai input ke perangkat lunak Minitab, persamaan relasional yang diperoleh diberikan dalam Persamaan 13.1.

$$Y(x) = A - B \tag{13.1}$$

$$A = \{(10.135 * r) + (37.291 * i) - (4.435 * t) + (12.642 * i)\}$$

$$B = \{(14.121 * r * r) + (1.823 * i * i) + (7.078 * t * t) + (0.238 * i * r) + (23.138 * r * i) + (3.938 * r * t) - (4.337 * r * i) - (22.867 * i * t) + (8.981 * i * r) + (7.324 * t * i) + 189.578$$

Sedangkan, 'r', 'i' dan 't' masing-masing mewakili parameter curah hujan, irigasi dan suhu, untuk periode waktu yang sesuai.

Persamaan yang diperoleh telah digunakan sebagai fungsi objektif atau persamaan kebugaran untuk algoritma kunang-kunang dengan tujuan untuk memaksimalkan nilai. Y(x) menunjukkan jumlah hasil panen jagung untuk periode waktu 'x' untuk lokasi tersebut. Nilai hasil panen yang diperoleh berada dalam kisaran 0 hingga 1 dan telah diubah menjadi format persentase hasil panen yang dapat dipahami pengguna.

Tahap keenam adalah penerapan algoritma kunang-kunang untuk mengidentifikasi tanaman optimal dan parameter iklim optimal untuk menghasilkan keuntungan maksimum. Tahap awal adalah menetapkan jumlah kunang-kunang dan membangun ruang pencarian yang akan digunakan dalam pencarian. Kinerja algoritma kunang-kunang dipengaruhi oleh faktor kontrol. Jadi jumlah kunang-kunang dapat dihitung menggunakan Persamaan 13.2.

$$nof = Cint\left(\sqrt[n]{no.of\ months\ x\ no.\ of\ control}\right) \tag{13.2}$$

Semua kunang-kunang virtual mencari solusi optimal berdasarkan pengetahuan kolektif yang diperoleh kunang-kunang yang lebih tua ke lokasi dan jalur terbaik. Kunang-kunang memilih jalur mereka berdasarkan daya tarik atau intensitas cahaya kunang-kunang lainnya. Daya tarik atau intensitas cahaya "I" kunang-kunang dapat dihitung menggunakan Persamaan 13.3.

$$\mathrm{I(r)} = (\mathrm{I_j}/\mathrm{r_{ij}}^{2}) \tag{13.3}$$

Sedangkan, 'r' adalah jarak antara kunang-kunang ke i dan ke j dan $\mathrm{I}_j = \mathrm{F(X)}$, fungsi objektif. Setiap kunang-kunang memiliki nilai koefisien daya tarik tertentu 'β' dan dapat dihitung dari Persamaan 13.4.

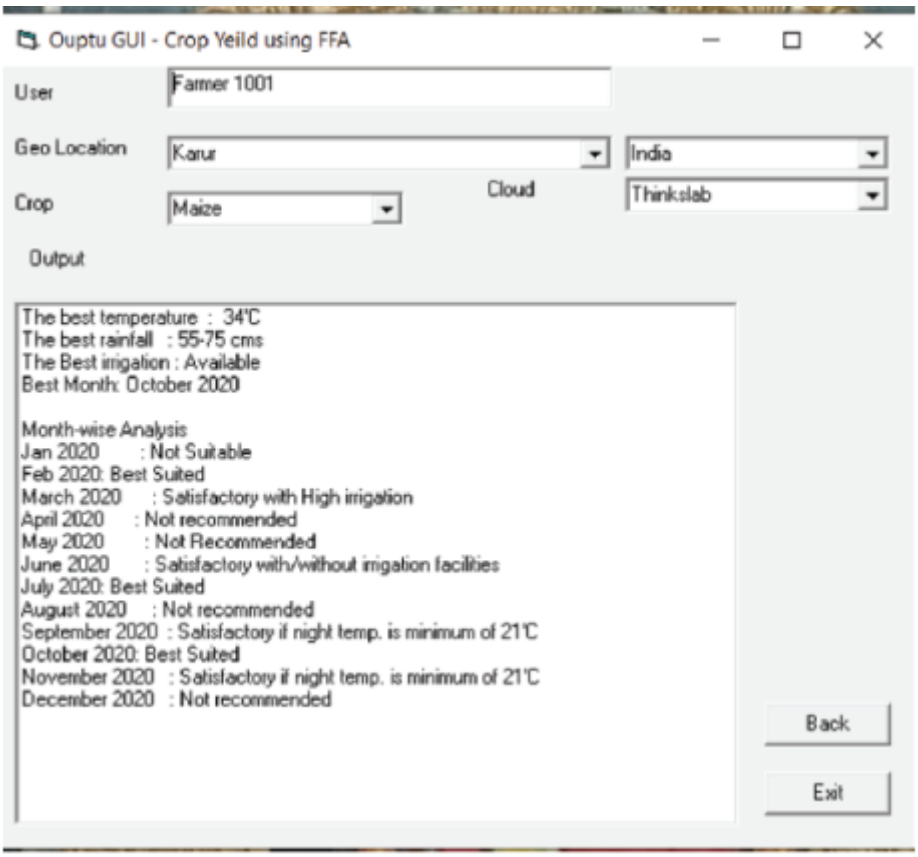
$$\beta = \beta_0 e^{-yr2} \tag{13.4}$$

Sedangkan, 'β_0' merupakan nilai daya tarik utama kunang-kunang dan 'γ' merupakan koefisien penyerapan cahaya. Pergerakan kunang-kunang dapat dihitung dengan menggunakan Persamaan 13.5.

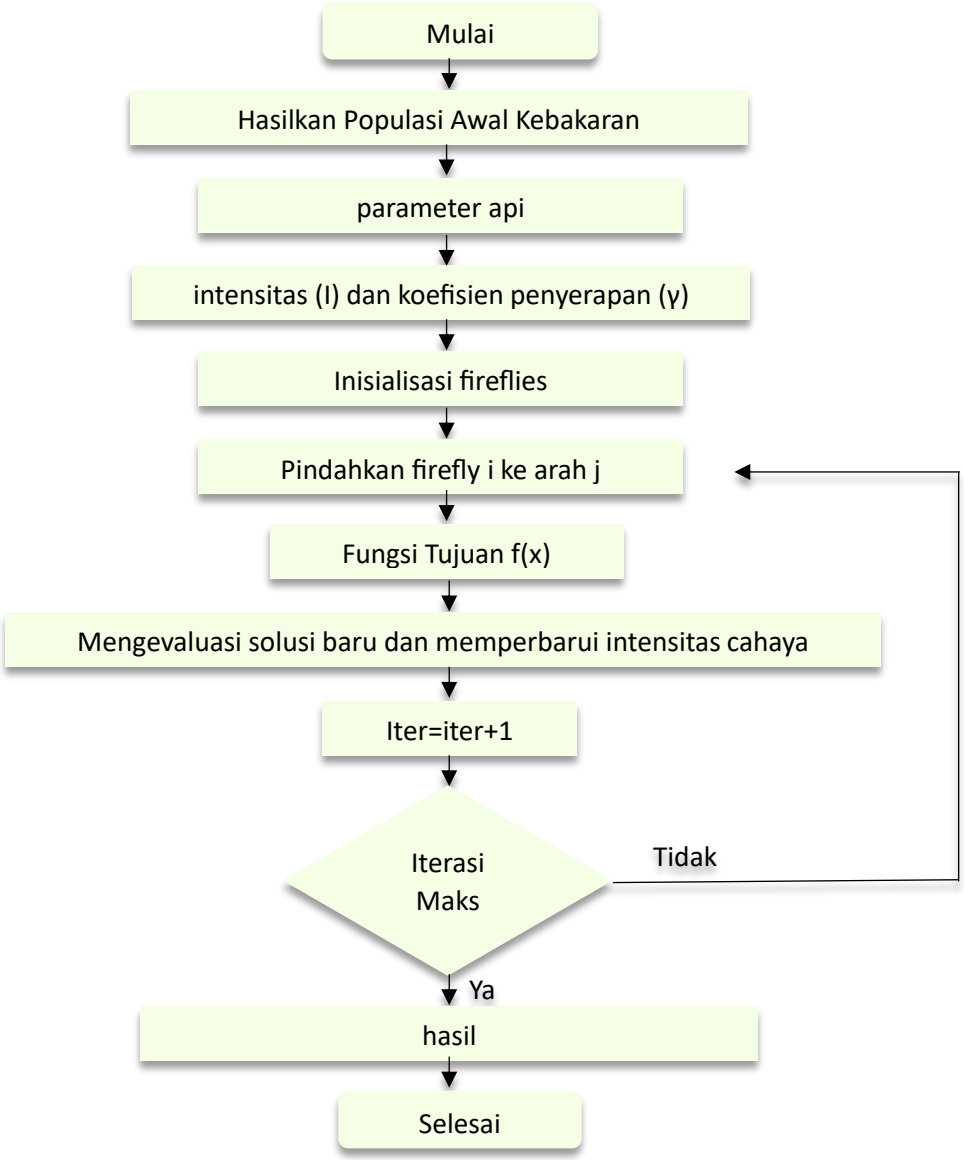
$$X_j = X_i + \beta_0 e^{-yr2} \left(X_j - X_i\right) + a\varepsilon^i \quad (13.5)$$

Padahal, hal itu terdiri dari posisi sebelumnya, daya tarik kunang-kunang dan parameter pengacakan untuk mencegah stagnasi. Jika posisi baru yang dihitung memiliki daya tarik yang lebih tinggi daripada yang sudah ada, maka kunang-kunang itu akan bergerak ke posisi baru yang diidentifikasi dan pergerakan harus berlanjut hingga simpul yang telah ditentukan tercapai, yaitu, tiga simpul dalam kasus ini.

Dalam setiap iterasi, jalur terbaik harus disimpan dan prosedur yang sama harus diulang hingga jumlah iterasi mencapai 100 yang ditunjukkan pada Gambar 13.3. Tahap ketujuh adalah validasi algoritma yang dikembangkan dan algoritma yang dikembangkan telah diuji untuk bulan Februari dan bulan Maret. Algoritma untuk kunang-kunang yang dikembangkan ditunjukkan pada Gambar 13.4.



Gambar 13.3 Output GUI dengan contoh output.

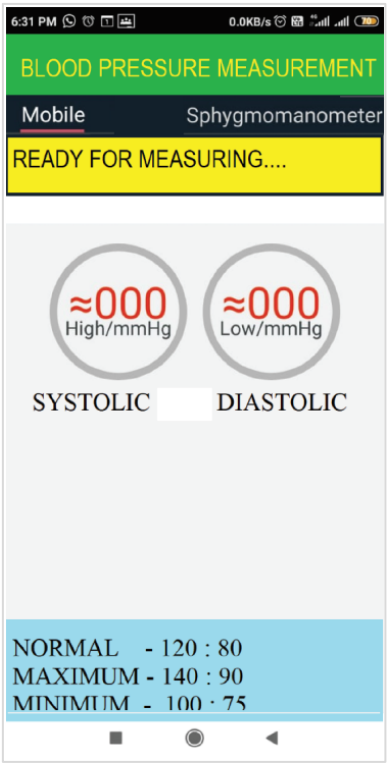


Gambar 13.4 Diagram alir untuk algoritma firefly.

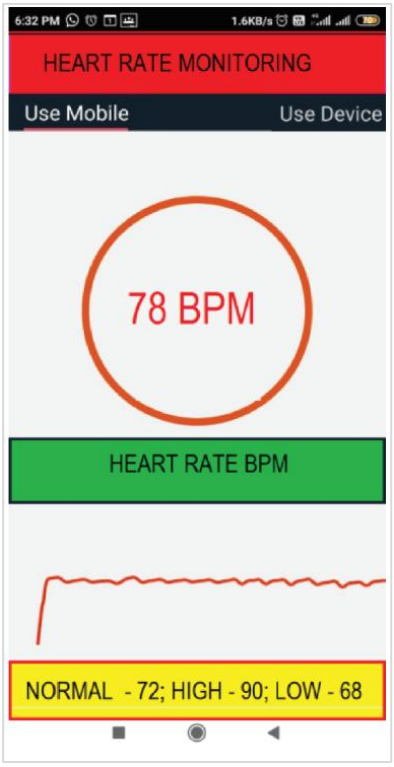
13.4 PEMANTAUAN LAYANAN KESEHATAN

Cloud yang dikembangkan juga telah digunakan untuk memantau layanan kesehatan petani dan modul yang dikembangkan juga menyarankan rekomendasi berdasarkan data yang diperoleh. Dalam pemantauan layanan kesehatan, tekanan darah petani telah dipantau, dan GUI aplikasi seluler Android ditunjukkan pada Gambar 13.5.

Detak jantung petani telah dipantau oleh sistem berbasis IoT yang dikembangkan dan GUI yang dikembangkan ditunjukkan pada Gambar 13.6.

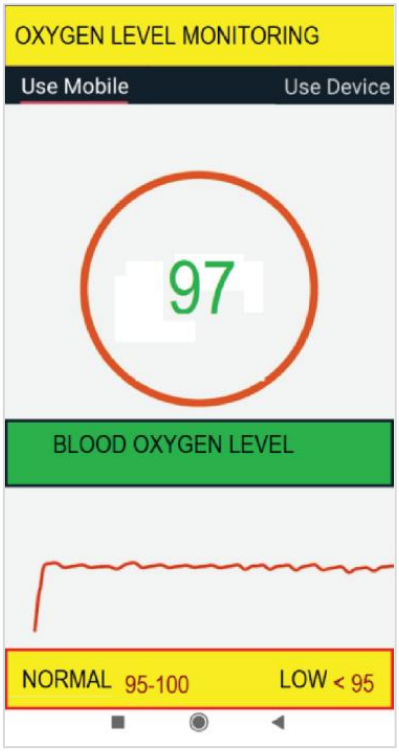


Gambar 13.5 GUI pengukuran tekanan darah.

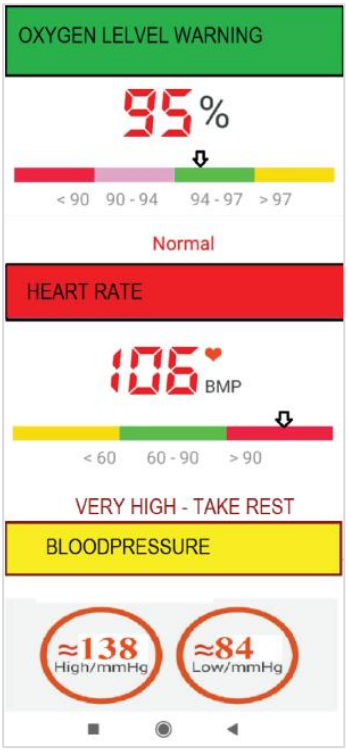


Gambar 13.6 GUI pengukuran denyut jantung.

Secara umum, petani tidak menjaga pola makan dan waktu, sehingga kadar oksigen harus dipantau dan GUI yang dikembangkan ditunjukkan pada Gambar 13.7.

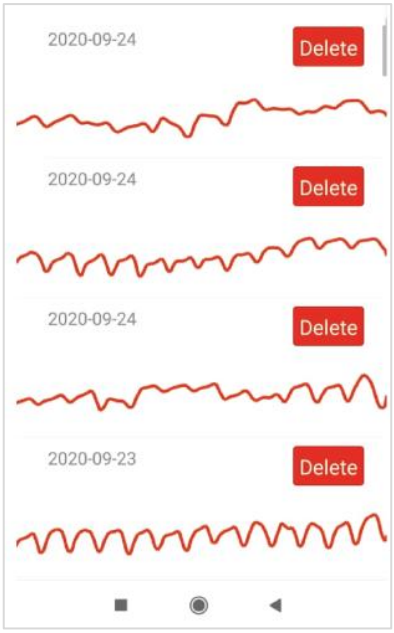


Gambar 13.5 GUI pengukuran tekanan darah.



Gambar 13.6 GUI pengukuran denyut jantung.

Data yang dikumpulkan dari petani telah dikirim ke cloud melalui sensor nirkabel yang sama yang digunakan untuk memantau kondisi pertanian. Dengan menganalisis data, cloud memberikan peringatan kepada petani untuk tindakan yang diperlukan; contoh pesan peringatan ditunjukkan pada Gambar 13.8. Modul yang dikembangkan juga menyimpan informasi yang dikumpulkan di cloud dan dapat ditafsirkan kapan saja. Contoh data yang disimpan ditunjukkan pada Gambar 13.9.



Gambar 13.9 Nilai basis data cloud.

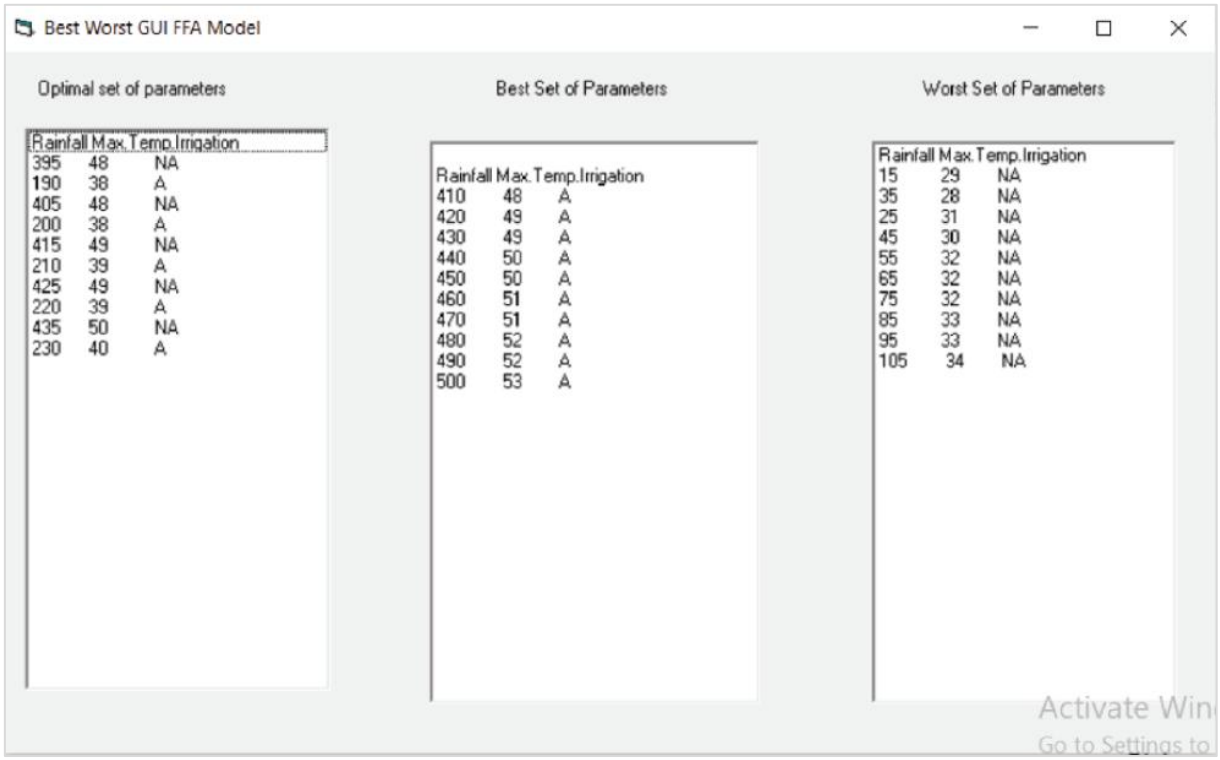
Dengan demikian, modul IOT yang dikembangkan telah digunakan untuk mengumpulkan data pertanian dan juga untuk memantau kesehatan petani. Dengan demikian, tujuannya bukan hanya untuk meningkatkan keuntungan budidaya tetapi juga kesehatan petani. Untuk pekerjaan eksperimental, cloud lab telah digunakan dan untuk penggunaan komersial, basis data cloud terpisah harus digunakan. Selain itu, data dapat dikirim ke pusat kesehatan primer terdekat dan ke kerabat untuk meditasi yang diperlukan; dengan demikian, kesehatan petani meningkat. Jenis aplikasi ini penting untuk daerah pedesaan di negara-negara berkembang, di mana para petani tidak memiliki banyak pengetahuan tentang kesehatan mereka.

13.5 HASIL DAN PEMBAHASAN

Wilayah pertanian jagung yang dipilih untuk studi eksperimental adalah desa Karur, negara bagian Tamilnadu, India selatan. Data yang dikumpulkan dari petani dan pelaku pertanian dipisahkan menjadi data primer dan sekunder. Data yang paling memengaruhi seperti hasil jagung, curah hujan, biaya input, fasilitas irigasi, dan suhu sebagai data primer dan data seperti musim panen, permintaan, pupuk, kualitas benih, dll., sebagai data sekunder.

Karena data dikumpulkan dalam rentang tahun yang sangat panjang, nilai rata-rata data dari 30 data petani jagung telah dipertimbangkan. Data yang bertentangan dan data yang tidak lengkap dihilangkan selama praproses. Dalam model algoritma kunang-kunang yang dikembangkan, input $\mathrm{I} = \{\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3\}$, sedangkan $\mathrm{X}_1 = \{\text{Curah hujan } \mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3, \ldots \mathrm{R}_n\}$ dalam

sentimeter, X2 = {suhu $T_1, T_2, T_3, \ldots T_n$} dalam celcius dan X3 = {Fasilitas irigasi tersedia, tidak tersedia} dalam Boolean. Fungsi hasil yang ditunjukkan dalam Persamaan 1 adalah fungsi maksimalisasi dan pada gilirannya keuntungan maksimum. Dengan mensimulasikan model kunang-kunang, hasilnya menunjukkan bahwa hasil jagung tertinggi adalah 68 kantong/hektar dan hasil terendah adalah 18 kantong/hektar untuk tahun 2020. Contoh keluaran dari model yang dikembangkan ditunjukkan pada Gambar 13.3.



Gambar 13.10 GUI untuk parameter optimal, terbaik, dan terburuk.

Setiap petani telah diberi ID unik dan petani harus menggunakannya untuk entri data dan untuk prediksi. Dalam mode geolokasi, petani diizinkan untuk memilih wilayah dan lokasi. Dalam modul ini, hanya satu lokasi, desa Karur yang dipertimbangkan. Akibatnya, model yang mapan ini membantu petani untuk meramalkan dan mengenali periode yang tepat untuk budidaya jagung secara maksimal dan ini pada gilirannya menghasilkan keuntungan yang tinggi. Model yang dikembangkan juga meramalkan parameter optimal, yang lebih baik untuk budidaya bagi petani. Dengan demikian, petani dapat melanjutkan dengan penilaian yang tepat untuk memaksimalkan pendapatan mereka. Bagian keluaran sampel ditentukan dalam Gambar 13.10. Sebanyak 1.103 fakta disurvei untuk 16 parameter dari tahun 2012 hingga 2019. Setelah praproses, tiga set data ditemukan lengkap dan memiliki pengaruh besar terhadap budidaya tanaman jagung dengan 288 titik data. Titik data dan parameter yang hilang dipertimbangkan untuk cakupan di masa mendatang.

Dalam Bab ini, kemungkinan penelitian baru tentang penggunaan algoritma kunang-kunang untuk mengidentifikasi dan memperkirakan parameter optimal untuk aplikasi hasil panen jagung dalam memaksimalkan keuntungan telah berhasil diimplementasikan menggunakan teknologi IOT. Tiga variabel iklim seperti suhu, curah hujan, dan fasilitas irigasi

telah dikumpulkan selama 96 bulan dan diproses terlebih dahulu untuk memastikan kebenarannya. Variabel tersebut telah digunakan untuk menghasilkan rumus relasional antara variabel menggunakan perangkat lunak Minitab dan telah digunakan sebagai fungsi kebugaran untuk algoritma kunang-kunang.

Hasil algoritma kunang-kunang telah divalidasi dengan data waktu nyata dan telah ditemukan bahwa hasilnya sesuai dengan data waktu yang ada. Dengan demikian, unit yang ditetapkan dapat digunakan untuk memprediksi parameter proses untuk tanaman lain dan untuk berbagai lokasi geografis. Keuntungan utama dari model yang dikembangkan adalah penyimpanan data berada di awan dan awan akan terus diperbarui untuk mendapatkan data yang akurat. Data awan juga dapat digunakan untuk statistik lain oleh pemerintah dan peneliti. Modul IOT yang dikembangkan juga berfungsi ganda untuk memantau kesehatan petani.

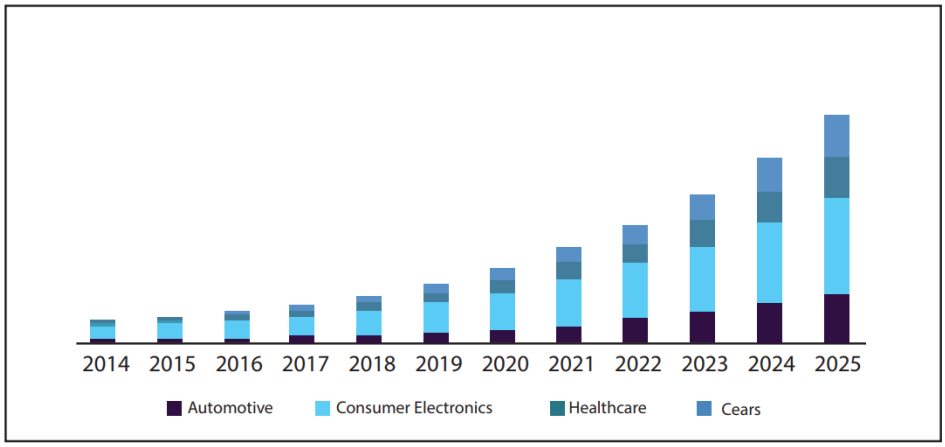
BAB 14
PENGENALAN GERAKAN BERBASIS PENGLIHATAN

Interaksi antara manusia dan mesin telah menjadi bagian dari layanan umum dan penting. Ini dikenal sebagai Antarmuka Manusia-Mesin dan dibagi menjadi deteksi, pelacakan, dan pengenalan. Dalam studi ini, pengenalan gerakan berbasis penglihatan telah dijelaskan secara rinci berdasarkan: Akuisisi, fitur, tingkat pengklasifikasi. Gerakan terdiri dari dua jenis: urutan statis dan dinamis, di sinilah teknik berbasis penglihatan memainkan peran penting.

Survei tentang studi penelitian tentang pendekatan pengenalan gerakan berbasis penglihatan telah dijelaskan secara singkat dalam makalah ini. Tantangan dalam semua perspektif pengenalan gerakan menggunakan gambar dirinci. Tinjauan sistematis telah dilakukan terhadap 100 makalah dan dipersempit menjadi 60 makalah dan dirangkum. Motif utama dari makalah ini adalah untuk memberikan landasan yang kuat pada pengenalan berbasis penglihatan dan menerapkannya untuk solusi di bidang medis dan teknik. Makalah ini menguraikan kesenjangan dan tren terkini untuk memotivasi para peneliti guna meningkatkan kontribusi mereka.

14.1 PENDAHULUAN

Isyarat adalah bentuk ekspresif manusia untuk menyampaikan informasi. Ketiga peran tersebut adalah Semiotik, Ergotik, dan Epistemik. Isyarat ditangkap melalui kamera atau sensor lalu diproses oleh komputer. Penelitian di bidang ini telah menjadi tren di pasar untuk aplikasi medis, lingkungan virtual, dll. Produk berbasis penglihatan memiliki variasi kamera dan fiturnya: lensa, bukaan, resolusi, rana, baterai.



Gambar 14.1 Pertumbuhan produk pengenalan gerakan di kawasan Asia-Pasifik

Gambar 14.1. menyajikan pertumbuhan produk ekonomi sistem pengenalan isyarat berbasis penglihatan di kawasan Asia-Pasifik yang diperoleh dari survei yang menunjukkan pertumbuhan luar biasa dalam produk pengenalan sejak 2014 dan berlanjut hingga 2025, karena mengamati keunggulan terapannya dalam pengenalan bahasa isyarat. Dalam bab ini disusun sebagai berikut: Bagian selanjutnya menjelaskan tantangan dalam pengenalan isyarat

berbasis penglihatan, Bagian III menjelaskan pemrosesan gambar, Bagian IV adalah tinjauan pustaka; Bagian V merangkum model yang ada dan Bagian VI menyajikan kesimpulan.
Isyarat telah diklasifikasikan berdasarkan berbagai kategori seperti berikut:

- Menunjuk
- Semaforik
- Pantomim
- Ikonik
- Manipulasi

14.2 MASALAH DALAM PENGENALAN GERAKAN BERBASIS PENGLIHATAN

Tantangan yang dihadapi oleh setiap perangkat pada pengenalan gerakan berbasis penglihatan dikategorikan dalam tiga unit dasar seperti pada Gambar 14.2, Gambar 14.3 dan Gambar 14.4. Tantangan tersebut dapat didasarkan pada parameter sistem komputerisasi, gerakan dan situasi lingkungan. Hal ini dapat menyebabkan perubahan dalam waktu respons, faktor biaya, efek iluminasi latar belakang seperti penskalaan dan rotasi, ukuran sampel kumpulan data, langkah pemrosesan gambar segmentasi dan ekstraksi fitur gerakan statis dan dinamis. Terlepas dari faktor-faktor ini, pemrosesan berbasis gambar waktu nyata telah dirancang dan dikembangkan untuk pengenalan gerakan.

Berdasarkan Gerakan

Gerakan berkisar dari yang sederhana hingga yang kompleks; statis ke dinamis. Berbagai masalah selama penangkapan gestur adalah: Penerjemahan, penskalaan rotasi gambar, multiview yang diperlukan untuk mendapatkan posisi gestur di semua arah: Pengenalan wajah (mengangkat alis, mengedipkan mata, lubang hidung, menggelengkan kepala, ekspresi), Pengklasifikasi bervariasi untuk jenis gestur Statis (NN) dan dinamis (DTW).

Berdasarkan Kinerja

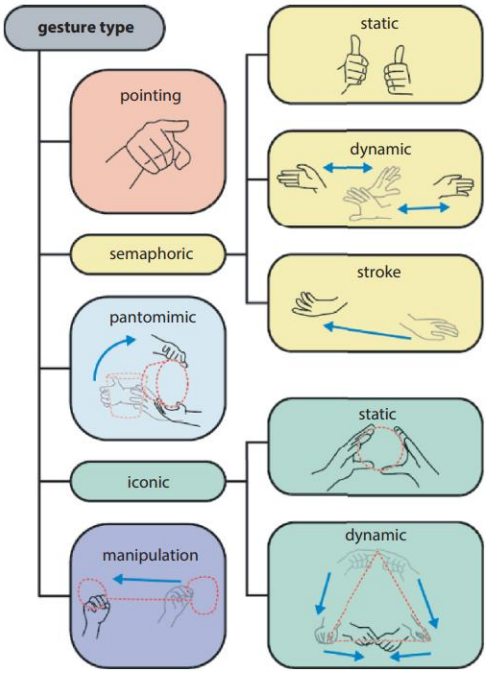
Perangkat yang dirancang untuk pengenalan berbasis Visi perlu memenuhi beberapa kriteria khusus: Waktu respons pemrosesan, waktu yang dihabiskan antara input ke output, Kecepatan transmisi dalam modul nirkabel, latensi yang dihindari, presentasi, efisiensi. Berapa banyak bingkai gambar per detik yang dapat ditangkap? Akurasi; Mempertahankan ketahanan tanpa mengubah latar belakang gambar. Ekstraksi fitur juga memperkirakan kinerja saat memilih fitur tingkat rendah seperti tepi dan histogram, yang menyebabkan tingkat kesalahan tinggi.

Berdasarkan Latar Belakang

Tahap praproses sangat penting karena tidak boleh menimbulkan gangguan dalam hal pencahayaan, oklusi, pencahayaan yang konsisten, jarak tetap dari lensa, dan objek yang tidak fokus. Persyaratan pengguna untuk mengenakan sarung tangan yang tidak praktis dan melepaskan aksesori.

14.3 PROSES LANGKAH DEMI LANGKAH DALAM PENGENALAN BERBASIS PENGLIHATAN

Gambar diklasifikasikan menjadi vektor dan digital. Gambar digital adalah larik bilangan riil 2 dimensi yang dibagi menjadi N-baris dan M-kolom. Perpotongan baris dan kolom tersebut dikenal sebagai piksel. Setiap piksel berwarna hitam dan putih yang biasanya berkisar antara 0 hingga 255. Dalam gambar berwarna, piksel dijelaskan berdasarkan jumlah RGB dengan empat tahap dalam pengenalan Berbasis Penglihatan.



Gambar 14.2 Klasifikasi gerakan.

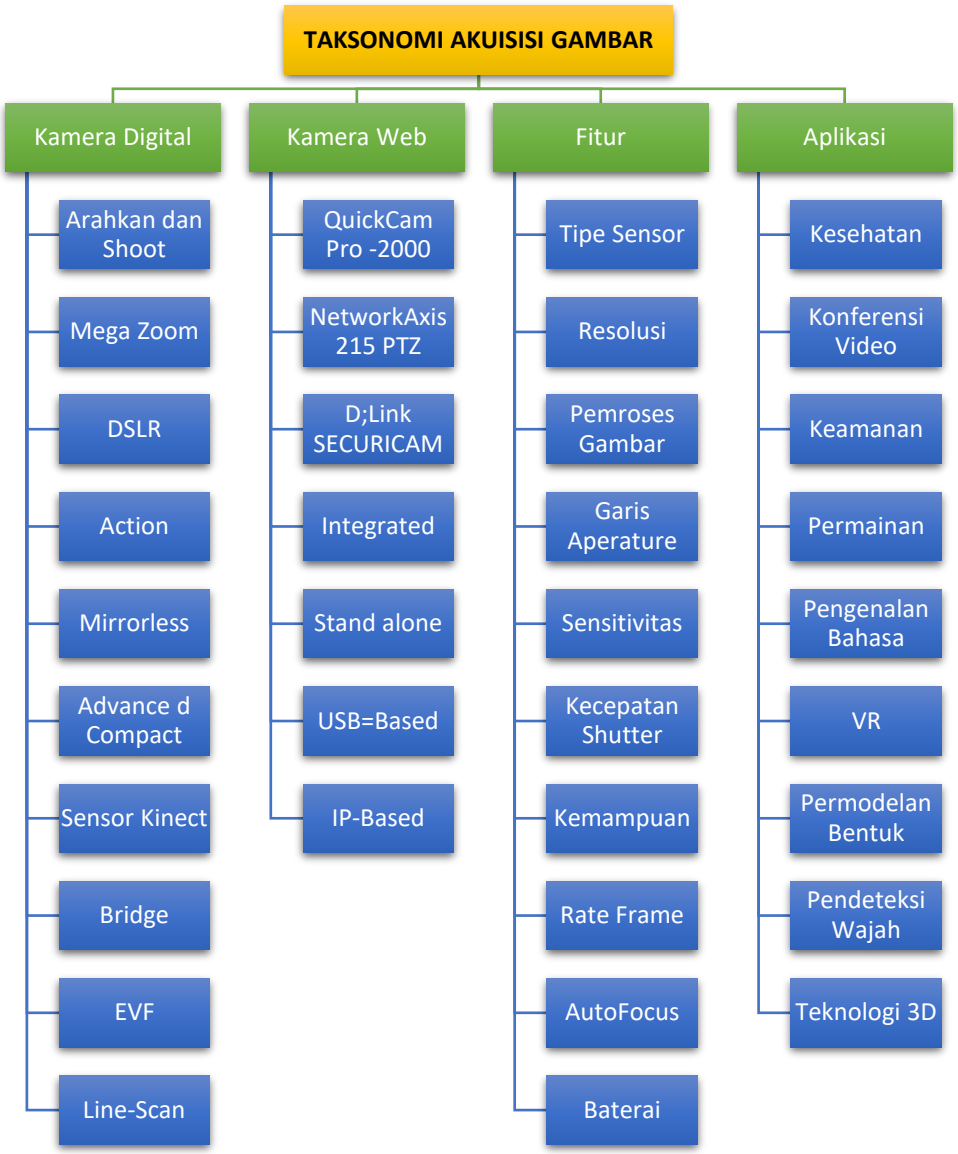
PENINGKATAN GAMBAR

Operasi titik	Operasi Spasial	Operasi Transformasi	Pewarnaan Semu
• Peregangan kontras • Pemotongan noise • Pemotongan jendela • Pemodelan histogram	• **Penghalusan derau** • **Penyaringan median**	• **Penyaringan linier** • **Penyaringan akar**	• **Pewarnaan palsu** • **Pewarnaan semu**

(a)

(b)

Gambar 14.3 (a) Operasi dalam Peningkatan Citra (b) Gerakan tangan dalam proses peningkatan citra tangan.



Gambar 14.4 Taksonomi akuisisi gambar.

Penginderaan

Kombinasi sumber "iluminasi" dan pantulan atau penyerapan energi dari sumber tersebut oleh elemen "pemandangan". 1) elemen penginderaan tunggal, 2) Sensor garis, 3) Sensor larik.

Praproses

Tujuannya adalah pengurangan derau, peningkatan dan pemulihan, segmentasi

- Salt & pepper – tersebar secara acak
- Gaussian – fluktuasi acak yang ditambahkan
- Speckle - nilai acak dikalikan
- Uniform – karena kuantisasi

a) Peningkatan citra

Ukuran manipulasi citra praproses untuk membuatnya dapat disesuaikan dengan aplikasi berbasis analisis citra lainnya. Proses ini mempertahankan karakteristik citra seperti tingkat derau yang dapat ditoleransi dan wilayah minat yang ditandai

b) Pemulihan citra
Ini meningkatkan citra dengan teknik degradasi yang dikenal. Citra yang salah ditampilkan dikembalikan untuk menyampaikan informasi aslinya. Transformasi ini dilakukan tanpa kehilangan konten citra karena merupakan proses yang objektif.

c) Segmentasi citra
Pembagian citra menjadi wilayah atau kategori, yang sesuai dengan objek atau bagian objek yang berbeda. Piksel gambar dimodifikasi berdasarkan dua parameter: diskontinuitas dan kesamaan di antara piksel gambar.

Ekstraksi Fitur

a) Deteksi fitur mengacu pada pencarian fitur dalam gambar, wilayah, batas.
b) Deskripsi fitur menetapkan atribut kuantitatif ke fitur detektor; misalnya, kita dapat mendeteksi sudut di batas wilayah, dan mendeskripsikan sudut tersebut berdasarkan orientasi dan lokasinya, yang keduanya merupakan atribut kuantitatif.
c) Metode pemrosesan fitur dibagi lagi menjadi tiga kategori utama tergantung pada apakah metode tersebut berlaku untuk batas, wilayah, atau seluruh gambar, parameter seperti skala, translasi, rotasi, iluminasi, dan sudut pandang.
d) Beberapa fitur berlaku untuk lebih dari satu kategori; deskriptor fitur harus setidak sensitif mungkin terhadap variasi parameter seperti skala, translasi, rotasi, iluminasi, dan sudut pandang.
e) Ekstraksi fitur adalah proses yang dengannya fitur tertentu yang menarik dalam suatu gambar dideteksi dan direpresentasikan melalui pemrosesan lebih lanjut. Ini adalah langkah penting dalam visi komputer dan solusi pemrosesan gambar yang menandai transisi dari representasi data bergambar ke non-bergambar.

14.4 KLASIFIKASI

Berbagai pengklasifikasi dalam pengenalan berbasis visi saat ini dirangkum dalam Tabel 14.1.

Tabel 14.1 Perbandingan pengklasifikasi.

Klasifikator	Manfaat	Keterbatasan
K-Nearest-Neighbour - Penggunaan kernel dan hyperplane untuk kategorisasi dataset 3D.	Efektif, Non-parametrik.	Pengaturan memori dan waktu memerlukan durasi yang lama, sehingga menghambat proses klasifikasi.
Jaringan Syaraf Tiruan - Node input yang berasal dari luar jaringan disebut node input dan hanya menyalin nilai.	Menghasilkan hasil yang baik dalam domain yang kompleks. Pengujian sangat cepat.	Proses pelatihan relatif lambat. Perlu meminimalkan Risiko Empiris.
Mesin Vektor Pendukung - Hiperbidang Pemisah Linier.	Data yang bersifat inheren lebih baik, tidak bergantung pada dimensi fitur.	Penyusunan parameter dan pemilihan kernel memerlukan perhatian khusus.

Pohon Keputusan - Jenis diagram alir di mana setiap simpul direpresentasikan sebagai pengujian.	Tidak membutuhkan pengetahuan awal. Mudah ditafsirkan.	Hanya menghasilkan satu keluaran dan sangat tergantung pada set data yang digunakan.
Klasifikasi Bayesian - Menentukan mean dan kovarians fungsi normal dan abnormal.	Proses komputasi menjadi lebih sederhana.	Untuk variabel dependen, hasil yang lebih akurat tidak dapat selalu dijamin.

Pemrosesan Citra Digital (DIP) adalah proses pengolahan citra digital menggunakan berbagai algoritma komputer. Makalah ini menyajikan tinjauan singkat dan tinjauan pustaka tentang teknik pemrosesan citra digital seperti pra-pemrosesan citra, kompresi citra, deteksi tepi, dan segmentasi. Desain dan Fabrikasi produk dengan tingkat pengenalan awal dan tinggi adalah Prototipe yang mampu mengenali bahasa isyarat secara otomatis untuk membantu orang tuli dan bisu berkomunikasi secara efektif. Algoritma convexity hull diimplementasikan untuk deteksi titik jari dan pengenalan angka. Perangkat lunak ini bertujuan untuk mengenali orientasi, pusat massa, dan berbagai fitur berbasis bentuk gerakan tangan. Menggunakan piksel 640*480 dengan fitur pengguna yang berbeda dapat membuat pengenalan menjadi lebih sulit. Dalam pendekatan lain Makalah ini mengilustrasikan pengenalan gerakan tangan dengan menggunakan perangkat android. Tujuan dari makalah ini adalah untuk mengenali 40 gambar yang berbeda. Karakteristik utamanya adalah menghitung centroid di tangan, keberadaan ibu jari dan jumlah puncak dalam gerakan tangan. Pengenalan didasarkan pada jaringan saraf tiruan. Elemen penginderaan perangkat android (aplikasi android Webcam) merasakan gerakan dan mengirimkannya sebagai input ke komputer. Gerakan tersebut dideteksi untuk tepinya.

Deteksi tepi adalah proses yang bertujuan mengidentifikasi titik-titik dalam gambar digital di mana kecerahan gambar berubah tajam atau memiliki diskontinuitas dengan tiga detektor Sobel, Canny, Prewitt. Penipisan adalah operasi morfologi yang digunakan untuk menghapus piksel latar depan yang dipilih dari gambar biner. Ini digunakan agar tepinya berada dalam garis tipis. Dalam metode ini detektor tepi yang jernih digunakan. Detektor menghitung gradien intensitas gambar di setiap titik, memberikan arah peningkatan terbesar yang mungkin dari terang ke gelap dan laju perubahan ke arah itu. Sudut A dan B adalah dua sudut yang diperlukan yang akan dimasukkan ke dalam lapisan jaringan saraf. Dengan kedua sudut tersebut, kita dapat secara tepat merepresentasikan arah hipotenusa dari titik P1 ke P2 yang merepresentasikan arah gambar tangan dengan rata-rata 77% beserta tingkat kebisingan yang lebih tinggi.

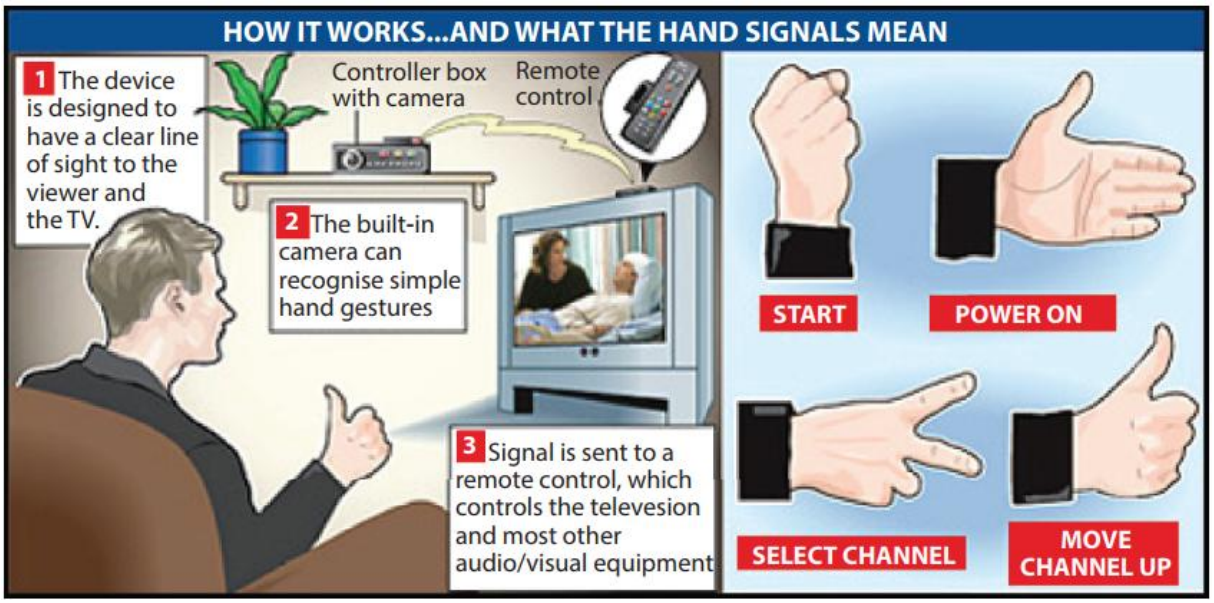
Peneliti mengusulkan metode pengenalan gerakan tangan statis yang efektif dengan menggunakan deskriptor efektif baru, yang disebut BCF berbasis DM yang mengodekan informasi bentuk tangan dari peta kedalaman dan merupakan gerakan dua tangan yang ringkas dan diskriminatif. Sistem ini terbatas pada oklusi jari dan inti tangan dan masih perlu dievaluasi dalam aplikasi nyata. Troli belanja robotik pintar meskipun ada beberapa input, ada fitur yang disebut pengenalan bahasa isyarat melalui perekaman video menggunakan kamera Kinect 2.0 yang memproses data dan memberikan output yang benar dalam bentuk rekaman suara dengan akurasi 95% dari gerakan dua tangan, dengan hanya satu kelemahan dari tingkat

kesalahan klasifikasi yang lebih tinggi.) mengusulkan metode pengenalan gerakan multimoda berdasarkan konvolusi 3-D dan LSTM konvolusional untuk pengenalan gerakan. CNN 3-D digunakan untuk mengekstraksi fitur spasiotemporal jangka pendek dari video input diikuti oleh LSTM konvolusional untuk mempelajari fitur spasiotemporal jangka panjang lebih lanjut.

Spatial Pyramid Pooling diadopsi untuk menormalkan fitur spasiotemporal. Penyetelan halus dievaluasi dan kinerja terkini pada dataset Isolde dan SKIG dilaporkan. Urutan gerakan dengan pencahayaan yang buruk tidak dapat dikenali dengan baik dalam Eksperimen; Gerakan yang cepat dan kecil sulit dikenali. mengenali Islip mantan bahasa isyarat India dikenali dan untuk memudahkan pemahaman, bahasa isyarat tersebut dimasukkan ke dalam sistem yang menghindari kebutuhan manual dasar untuk mengajarkan bahasa isyarat, tetapi estimasi nilai eigen menghabiskan waktu. Dalam yang terakhir, 24 alfabet bahasa isyarat India dalam video langsung dapat dikenali, dengan tingkat pengenalan yang rendah. Dalam sistem pengenalan gerakan tangan dengan pembelajaran mendalam yang disempurnakan menggunakan jaringan saraf konvolusi (CNN) dan kamera kedalaman untuk interaksi manusia-komputer. Ia juga menggunakan sensor Kinect untuk mendeteksi gerakan tangan dengan akurasi 84,67% untuk Ejaan Jari ASL dengan gangguan besar dari tekstur dan oklusi. Dalam makalah ini tiga representasi kedalaman berbasis gambar sebagai Kedalaman Dinamis.

Dynamic Depth Image (DDI), *Dynamic Depth Normal Image* (DDNI) dan *Dynamic Depth Motion Normal Image* (DDMNI) diusulkan, di mana DDI terutama mengeksploitasi dinamika postur, dan DDNI dan DDMNI, yang dibangun di atas vektor norma, secara efektif mengeksploitasi struktur 3D, dibatasi oleh objek yang terlibat dalam tindakan dengan pola gerakan serupa yang sulit dibedakan dalam informasi peta kedalaman yang ditangkap oleh peta kedalaman. Dua tingkat pemrosesan gambar melalui antarmuka manusia-komputer; satu adalah dengan fitur seperti Haar yang mewakili fitur seperti Haar lebih stabil dalam menyediakan objek warna kulit yang benar dan dengan mudah membedakan daerah gelap dari kulit. Yang lainnya adalah ada-booster yang membantu dalam menyediakan nilai biner yang akan dibaca oleh komputer menggunakan fitur Haar. ASL mengusulkan arsitektur air terjun untuk menggabungkan submodul. Manusia berinteraksi dengan aksesori, gerakan yang secara langsung menangani komputer.

Gyro MPU6050 memperoleh gerakan dan pengenalan pra-perlakuan untuk bereksperimen menggunakan sensor sikap yang diikatkan di pergelangan tangan orang tersebut. Tinjauan serupa yang dilakukan pada berbasis penglihatan dilakukan oleh Carlos dan Robin, di mana beberapa kamera kedalaman seperti Microsoft Kinect, ASUS Xtion, Mesa Swiss Ranger dan kamera video Stereo untuk identifikasi gerakan yang lebih baik. Ada dua tahap: Lokalisasi tangan dan Lokalisasi gerakan. Lokalisasi tangan didasarkan pada visi komputer sedangkan Lokalisasi gerakan didasarkan pada pembelajaran mesin. Keterbatasan penggunaan HMM memerlukan penambahan model bigram dan sistem tidak dapat mengenali kosakata yang lebih besar. Dalam satu pendekatan metodologi ASL, survei, analisis hasil, kekurangan diambil dalam. Juga satu aplikasi berbasis penglihatan pribadi yang khas telah diilustrasikan pada Gambar 14.5. Subha dan Balakrishnan telah merancang 32 posisi atas dan bawah dari gerakan satu tangan dengan gambar. Sekarang 32 tanda ini sedang dibandingkan dengan 320 gambar yang dilatih. Setelah proses ekstraksi dan ambang batas selesai, selanjutnya dipindahkan ke fase pengujian menggunakan perangkat lunak apa pun, katakanlah MATLAB.



Gambar 14.5 Remote yang dikendalikan penglihatan (www.zdnet.com)

Selain itu, data mentah yang diperoleh melalui metode visual (kamera) perlu dikirimkan ke banyak pengguna, perangkat penyimpanan, dan aplikasi tertentu. Transmisi ini dapat dilakukan secara lebih global dengan penggunaan komputasi awan. Ini memungkinkan gambar visual yang ditangkap untuk diproses menjadi informasi yang diperlukan, dengan kapasitas hingga satu tetra byte. Dengan mengakses data mentah menggunakan server awan, ini memberikan beberapa manfaat dari perspektif pengguna: lebih banyak pengguna dengan biaya terbatas, beberapa perangkat dapat dihubungkan secara bersamaan, tidak ada persyaratan perangkat keras khusus, dan akses data yang lebih cepat. Namun, hal ini juga terbatas karena alasan seperti tidak tersedianya pemulihan data dan tidak dapat mengendalikan pengoperasian data yang ada di cloud.

Pengenalan isyarat yang dirancang adalah untuk mempersempit kesenjangan antara penyandang disabilitas dan orang normal. Dua metode diusulkan: Pendekatan Penglihatan dan Sensor. Informasi dan penelitian yang sama banyaknya telah dilakukan untuk keduanya. Dalam makalah ini, berbagai macam pekerjaan Berbasis Penglihatan, beserta masalah, batasan, dan tingkat keberhasilannya dipertimbangkan. Model sebelumnya dibandingkan dengan berbasis penglihatan yang ada di bidang aplikasi: pengenalan isyarat untuk kontrol robotik, pengenalan bahasa isyarat, pemrosesan gambar. Pekerjaan di masa mendatang menghasilkan peluang untuk merancang perangkat yang efisien, ringkas, dan layak untuk pengenalan isyarat berdasarkan gambar. Mengatasi masalah yang disebutkan pasti akan terbukti bermanfaat bagi orang-orang yang mengalami gangguan bicara dan pendengaran.

BAB 15
PEMFILTERAN SPAM MENGGUNAKAN KECERDASAN BUATAN

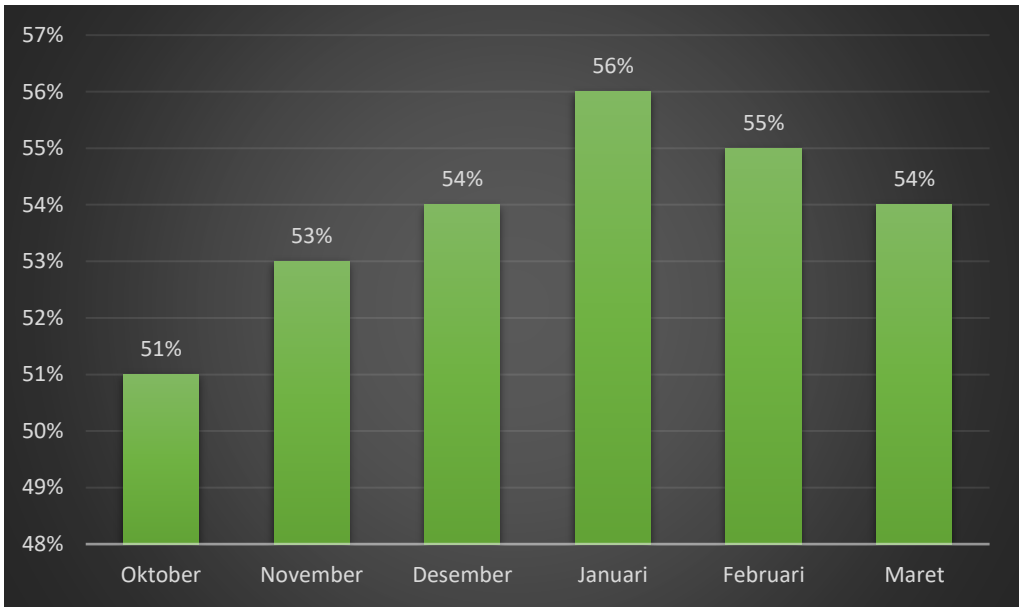
Selama kita semua mengelola masalah email yang tidak diminta, mungkin ada kebutuhan mendesak untuk pengembangan filter antispam yang kuat dan teguh. Saat ini, pembelajaran gawai, yang merupakan bagian dari kecerdasan buatan, mulai digunakan untuk menemukan dan menentukan filter antispam. Dalam makalah ini, kami membahas mekanisme dan algoritme penyaringan spam email elektronik berbasis pembelajaran sistem. Makalah ini akan membahas berbagai pemikiran, upaya, efisiensi, dan berbagai studi tentang tren dalam penyaringan email sampah. Sejarah menjelaskan paket strategi pembelajaran gawai untuk membersihkan email antispam dari operator layanan email utama seperti Gmail, Yahoo, Outlook, dan sebagainya. Kami akan membahas teknik penyaringan email yang tidak diminta dan berbagai upaya yang dilakukan oleh berbagai peneliti dalam memerangi email yang tidak diminta melalui strategi penguasaan gawai. Di sini, kami membuat perbandingan kekuatan dan kelemahan algoritma dan teknik pembelajaran mesin yang sudah ada dan berbagai masalah studi terbuka dalam penyaringan spam. Kami mungkin menyarankan untuk mendapatkan pengetahuan yang mendalam dan juga pembelajaran yang mendalam tentang musuh karena teknologi ini akan membuat kita mampu menangani ancaman email spam secara efektif.

15.1 PENDAHULUAN

Email massal yang dikenal sebagai surat yang tidak diminta telah berkembang menjadi gangguan besar di internet saat ini. Seorang spammer adalah seseorang yang mengirim email tidak sah ini. Mereka memperoleh alamat email dari berbagai situs web, sistem media sosial, ruang obrolan, survei, birokrasi kontak, dan sebagainya. Email surat yang tidak diminta bertanggung jawab atas pemborosan waktu seseorang, kapasitas penyimpanan sistem, dan bandwidth komunitas. Volume besar email sampah yang masuk ke komunitas komputer memiliki dampak buruk yang ekstrem pada memori server, bandwidth komunikasi, daya CPU, dan waktu pengguna. Lebih dari 77% email di seluruh dunia adalah email sampah, dan email sampah ini meningkat setiap tahun. Hal itu sangat menjengkelkan bagi pengguna yang menerima email yang tidak mereka minta, yang dikenal sebagai email sampah.

Hal terburuknya adalah banyak pelanggan yang terjebak oleh spammer dan peretas yang mengirim email sampah tersebut untuk memikat pelanggan agar jatuh ke dalam penipuan internet dan praktik penipuan spammer. Mereka berpura-pura berasal dari perusahaan tepercaya tetapi niat mereka adalah membujuk orang untuk mengungkapkan informasi non-publik mereka seperti kata sandi, info rekening bank, nomor kartu kredit, dan sebagainya. Pada tahun 2018, survei zona ke-4 menunjukkan bahwa email spam merupakan 51% dari keseluruhan email. Dampak spam karena penggunaan aset yang tidak produktif pada protokol pengalihan surat sederhana (SMTP) karena mereka menginginkan metode untuk mengirim sejumlah besar pesan yang tidak diperlukan yang dikenal sebagai surat yang tidak diminta. Pada Gambar 15.1, kami menunjukkan luasnya pesan surat yang tidak diminta yang berisi virus dan kode malware pada tahun 2018 dan 2019 sebagai berikut:

Perusahaan layanan Email terkemuka seperti Yahoo, Outlook, dan Gmail telah menggunakan berbagai kombinasi gawai untuk mempelajari teknik seperti jaringan saraf dalam filter surat yang tidak diminta untuk menangani ancaman yang ditimbulkan oleh spam email secara efisien. Dari sistem komputer koleksi besar, teknik pembelajaran mesin tersebut memiliki kapasitas untuk mempelajari dan mengidentifikasi email spam dan pesan yang tidak diperlukan dengan cara menganalisis pesan yang sebanding. Penggunaan peraturan yang berlaku sebelumnya, teknik pembelajaran gawai beradaptasi dengan berbagai kondisi dan melakukan lebih dari sekadar memeriksa email sampah Gmail dan Yahoo. Model pembelajaran gawai ini membuat standar baru sendiri tergantung pada apa yang telah dipelajarinya selama operasi penyaringan surat sampah ini.



Gambar 15.1 Kapasitas email spam zona 4 tahun 2018 hingga area 1 tahun 2019.

Google memiliki model pembelajaran mesin paling unggul yang dapat menemukan dan menyaring email sampah dan phishing dengan presisi 99%. Hasilnya adalah hanya satu dari seribu pesan yang akan berhasil lolos dari saluran email sampah. Sejalan dengan statistik Google, 50 hingga 70% email yang dapat diterima melalui Gmail adalah email spam. Alat Google seperti penjelajahan aman untuk mengklasifikasikan situs web yang memiliki URL berbahaya diintegrasikan melalui model deteksi Google. Kinerja deteksi phishing Google secara keseluruhan menjadi lebih efektif melalui pengenalan perangkat yang menangguhkan pengiriman email dan pesan untuk beberapa waktu guna melakukan pemeriksaan lebih ketat terhadap pesan phishing karena pesan tersebut relatif mudah ditemukan saat dianalisis secara massal. Pesan phishing dan email email sampah ini sengaja diperkenalkan terlambat untuk melakukan pemeriksaan lebih mendalam terhadap email mencurigakan tersebut pada saat yang sama ketika pesan lain diperkenalkan tepat waktu dan algoritme diperbarui secara waktu nyata. Email yang dapat ditangani dengan penundaan ini jumlahnya hanya 0,05%.

Berikut ini adalah kategori khusus teknik penyaringan email sampah yang banyak digunakan untuk mengatasi masalah email yang tidak diminta:

1. Pendekatan Penyaringan Email Sampah Berbasis Konten: Pendekatan ini menganalisis frasa, kemunculannya, dan distribusi kata dan frasa dalam bingkai email, lalu

menggunakannya untuk menyaring email masuk yang tidak diminta. Pendekatan ini digunakan untuk membuat panduan penyaringan otomatis dan mengkategorikan email dengan menggunakan algoritme penguasaan sistem seperti klasifikasi Naïve Bayes, SVM, ok-nearest neighbor, dan jaringan saraf. Algoritme tersebut akan disebutkan di bagian selanjutnya.

2. Metode Penyaringan Email Sampah Berbasis Kasus: Ini adalah strategi penyaringan spam yang paling populer di mana setiap email sampah dan email non-spam ditarik keluar dari email seseorang dengan menggunakan model pengumpulan. Kemudian langkah pra-pemrosesan dilakukan untuk mengubah email menggunakan antarmuka pelanggan, ekstraksi karakteristik, pilihan, pengelompokan fakta email, dan pemeriksaan metode. Kemudian fakta yang diperiksa dibagi menjadi beberapa unit vektor. Cepat atau lambat, mesin akan mengetahui teknik yang digunakan untuk mendidik kumpulan data dan menguji catatan untuk memutuskan apakah email yang masuk adalah email sampah atau bukan.
3. Pendekatan Penyaringan Email Sampah Berbasis Aturan atau Heuristik: Metode ini memeriksa berbagai macam gaya yang mungkin merupakan ekspresi umum terhadap pesan yang dipilih menggunakan pedoman saat ini. Peringkat pesan meningkat dengan pola yang serupa, atau dihapus untuk peringkat jika ada gaya yang tidak sesuai. Pesan dapat dianggap sebagai email sampah, jika pesan melebihi ambang batas unik, dalam kasus lain dapat dianggap sebagai email yang tidak diminta. Teknik-teknik tersebut perlu diperbarui selama bertahun-tahun, yang merupakan cara yang baik untuk mengatasi spammer yang terus-menerus memperkenalkan pesan email sampah baru yang dapat dengan mudah lolos dari pengenalan sebagai email yang tidak diminta. Contoh pembersihan email sampah berbasis aturan adalah pembunuh spam.
4. Pendekatan Penyaringan Spam Berbasis Kemiripan Sebelumnya: Pendekatan ini menggunakan teknik AI berbasis memori, atau berbasis contoh, untuk mengurutkan pesan masuk berdasarkan kedekatannya dengan pesan yang disimpan. Lokasi email digunakan untuk membuat vektor ruang multidimensi, yang diterapkan untuk memetakan kasus baru sebagai fokus. Kasus baru ini kemudian dialokasikan ke kelas penyelesaian paling jauh dari kasus persiapan terdekat-k. Teknik ini menggunakan k-tetangga terdekat (KNN) untuk membersihkan pesan spam.
5. Teknik Penyaringan Email Tak Diminta Adaptif: Pendekatan ini merasakan dan menyaring email tak diminta dengan mengelompokkannya ke dalam kelas yang berbeda. Pendekatan ini membagi bingkai email ke dalam berbagai perusahaan; setiap perusahaan memiliki teks simbolis.

Penilaian dilakukan antara setiap email masuk dan setiap perusahaan, lalu persentase kemiripan dihitung untuk memilih perusahaan tempat email tersebut berada.

15.2 ARSITEKTUR SERVER EMAIL DAN TAHAPAN PEMROSESAN EMAIL

Arsitektur - Penyaringan Spam Email

Penyaringan email sampah mengacu pada pengurangan email yang tidak diminta hingga seminimal mungkin. Dalam penyaringan email, email diproses dan disusun ulang sesuai dengan beberapa persyaratan positif. Filter email digunakan untuk mengontrol hal-hal berikut, termasuk email masuk, membersihkan email sampah, menemukan dan menghapus email

yang mengandung kode berbahaya yang mencakup virus dan malware. Protokol SMTP bertanggung jawab atas jalannya email.

Vendor penyedia internet luar biasa mengatur saluran spam di setiap lapisan kerangka kerja, sebelum pakar email, atau di jalur email yang dekat dengan firewall. Berdasarkan kebijakan perlindungan yang ditentukan, firewall, yang merupakan sistem keamanan jaringan, mengelola dan menampilkan video pengunjung jaringan yang masuk dan keluar. Server email membantu dalam menerapkan anti-email yang tidak diminta dan solusi anti-virus memungkinkan dalam menyajikan langkah-langkah keamanan lengkap untuk email di seluruh komunitas. Add-on disiapkan di antara gadget titik akhir untuk berfungsi sebagai perantara tempat filter dapat dijalankan di klien.

Cara Kerja Filter Spam Email Gmail, Yahoo, dan Outlook :

Gmail, Outlook.com, dan Yahoo secara aktif menjalankan berbagai formula penyaringan spam untuk mengirimkan email resmi yang paling efektif kepada pengguna dan membersihkan email yang sah atau sampah. Di sisi lain, formula tersebut terkadang secara keliru memblokir pesan yang dapat diandalkan. Laporan tersebut mengatakan, sekitar 23% email yang relevan umumnya gagal mencapai kotak masuk penerima yang sebenarnya. Berbagai mekanisme dirancang melalui penyedia layanan email untuk menggunakan filter anti-spam guna membatasi ancaman yang dimodelkan melalui malware, phishing, dan banyak lainnya bagi pengguna email. Banyak mekanisme yang digunakan untuk menentukan tingkat bahaya untuk setiap email yang masuk. Ini termasuk batasan spam tingkat pertama, kerangka kebijakan pengirim, daftar putih dan daftar hitam, dan peralatan verifikasi penerima. Mekanisme tersebut dirancang untuk digunakan oleh satu atau beberapa pengguna.

Filter Spam - Gmail

Ratusan aturan digunakan oleh pusat data Google untuk menilai keaslian email. Bergantung pada kemungkinan apakah fitur tersebut adalah spam atau bukan, setiap aturan menjelaskan fitur email sebagai email sampah dan biaya statistik pasti yang menyertainya. Kemudian dengan menggunakan bobot kepentingan ini, konstruksi persamaan dilakukan. Peringkat digunakan untuk melawan ambang sensitivitas guna melakukan pengujian yang diputuskan melalui pembersihan email sampah pengguna. Oleh karena itu, email tersebut diberi label sebagai email yang tidak diminta atau email yang valid.

Untuk jenis email tersebut, Google menggunakan algoritme pembelajaran perangkat pengenalan spam seperti regresi logistik dan jaringan saraf. Google juga mempraktikkan Optical Individual Popularity (OCR) untuk melindungi spam gambar bentuk pelanggannya. Algoritme pembelajaran perangkat dirancang dan dikembangkan untuk mencampur dan memberi peringkat pada kumpulan besar hasil pencarian mesin pencari sebagai akibatnya, pelanggan diizinkan untuk menghubungkan berbagai faktor Gmail guna menyederhanakan kategori email sampah. Popularitas area dan tajuk tautan merupakan elemen penting yang bertanggung jawab atas evolusi surat yang tidak diminta dari waktu ke waktu, karena pesan-pesan ini tiba-tiba masuk ke dalam folder surat sampah.

Filter Email Spam - Yahoo

Yahoo adalah penyedia layanan webmail tanpa batas terbesar dengan 320 juta klien di seluruh dunia. Yahoo memiliki kalkulasi sendiri yang digunakan untuk mengidentifikasi pesan spam. Penyaringan URL, konten email, dan penolakan spam dari klien adalah teknik dasar yang digunakan Yahoo untuk mengenali pesan spam. Yahoo memfilter email melalui area yang mirip

dengan Gmail, yang memfilter email menggunakan alamat IP. Yahoo memiliki mekanisme sendiri untuk mencegah pelanggan yang sah agar tidak salah diidentifikasi sebagai spammer. Tidak seperti daftar hitam, Yahoo memberikan daftar putih internal dan sertifikasi pengembalian dengan bantuan yang memungkinkan pengguna menentukan daftar penerima dan pengirim.

Filter spam ini memungkinkan pengguna untuk menggunakan campuran daftar putih dan fitur penanggulangan email tak diminta lainnya sebagai cara untuk mengurangi jumlah pesan sah yang dapat secara keliru dikategorikan sebagai email tak diminta. Namun, penggunaan daftar putih akan membuat filter menjadi sangat ketat dan implikasinya adalah bahwa setiap konsumen yang tidak disetujui akan diblokir secara otomatis. Banyak sistem anti-spam menggunakan daftar putih otomatis. Dalam contoh ini, email pengirim yang tidak disebutkan namanya diperiksa terhadap basis data; jika tidak ada catatan spam, pesan mereka dikirim ke kotak masuk penerima.

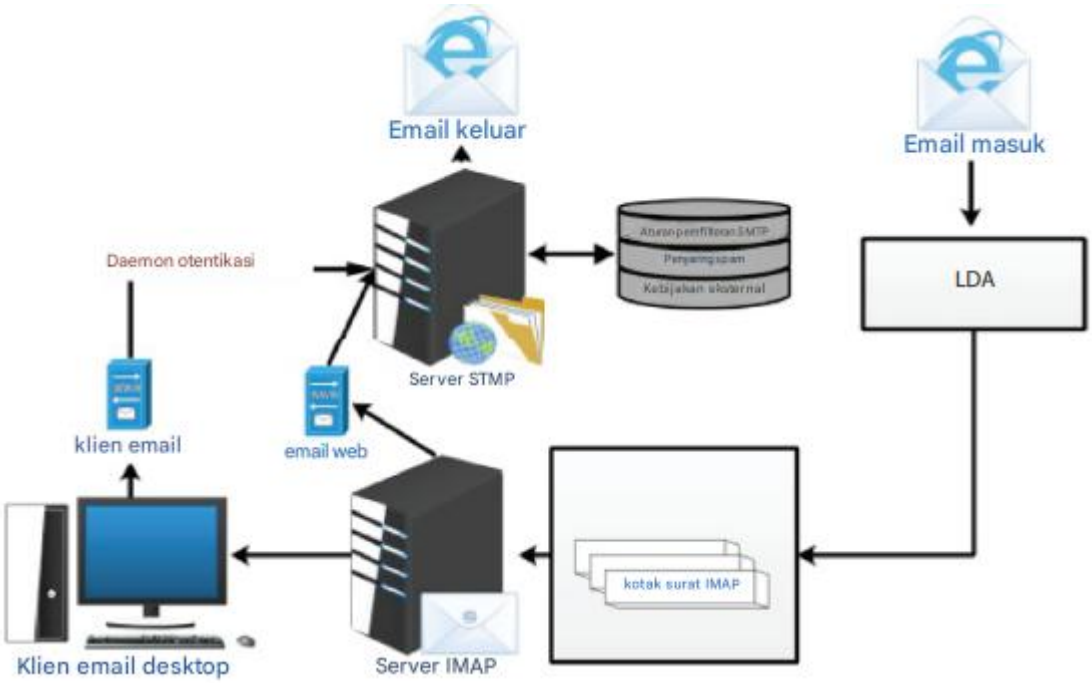
Filter Spam Email - Outlook

Outlook.com diubah menjadi bahasa rencana metro Microsoft dan secara langsung menduplikasi antarmuka Microsoft Outlook. Header email disertakan bidang-bidang, misalnya, wilayah pengirim, lokasi penerima, atau stempel waktu yang menunjukkan saat pesan diubah melalui pekerja transisi ke dealer pengiriman pesan yang bertindak sebagai kantor untuk mengatur pengiriman. Baris header, untuk sebagian besar komponen, dimulai dengan "dari" dan mengalami beberapa penyesuaian pada setiap faktor pergerakannya dimulai dengan satu karyawan kemudian ke yang berikutnya melalui pekerja perantara. Header memungkinkan konsumen untuk melihat rute yang dilalui email, dan waktu yang dibutuhkan oleh setiap karyawan untuk menangani email. Fakta yang dapat diakses perlu melalui beberapa penanganan sebelum pengklasifikasi dapat menggunakannya untuk memisahkan.

Penyaringan Spam Email - Proses

Pesan Email dihasilkan dari komponen besar yaitu header dan frame. Header adalah wilayah yang memiliki statistik luas tentang substansi email. Ini mencakup masalah, pengirim, dan penerima. Frame adalah bagian tengah email. Ini dapat berisi informasi yang tidak memiliki data yang telah ditentukan sebelumnya. Model terdiri dari halaman situs web, suara, video, informasi sederhana, gambar, dokumen, dan markup HTML. Header email adalah bidang yang dilindungi, misalnya, lokasi pengirim, wilayah penerima, atau stempel waktu yang ditampilkan saat pesan dikirim oleh orang yang bertransisi ke penjual pengiriman pesan yang bertindak sebagai tempat kerja untuk mengatur pengiriman.

Baris header untuk sebagian besar dimulai dengan "dari" dan mempelajari beberapa penyesuaian pada suatu titik yang bergerak mulai dari satu pekerja kemudian ke pekerja berikutnya melalui pekerja di dalam pusat. Header memungkinkan konsumen untuk melihat jalur email yang dilalui, dan waktu yang dibutuhkan oleh setiap pekerja untuk menangani email. Statistik yang dapat diakses ingin menjalani beberapa penanganan sebelum pengklasifikasi dapat menggunakannya untuk mengisolasi yang ditunjukkan pada Gambar 15.2.



Gambar 15.2 Representasi struktur server email dan proses penyaringan spam.

Pra-Penanganan

Ini adalah tahap paling awal yang dilakukan saat email yang akan dikirim diterima. Tahap ini melibatkan tokenisasi.

Perpajakan

Metode ini menghapus kata-kata dalam bingkai email. Metode ini juga mengubah pesan menjadi elemen-elemen utamanya. Metode ini mengambil email dan membaginya menjadi serangkaian gambar representatif yang disebut token. Gambar representatif ini diambil dari bingkai email, tajuk, dan subjek. Beberapa ilmuwan menekankan bahwa semakin dekat dengan penggantian informasi dengan gambar atribut yang tepat akan membuang semua kualitas dan frasa dari email yang terbatas pada pertimbangan makna.

Pemilihan Fitur

Langkah selanjutnya dari tahap pra-penanganan adalah tahap pemilihan elemen. Sorot preferensi jenis yang lebih rendah dalam proporsi inklusi spasial yang secara layak terkenal sebagai elemen menarik dari pesan email sebagai vektor elemen yang dikemas. Teknik ini menguntungkan sementara ukuran pesannya sangat besar dan penggambaran yang diringkas diperkirakan akan membuat tantangan pencocokan teks atau foto menjadi cerdas.

Tingkatkan pemerasan harga, termasuk warisan, lotere, visa dan petunjuk kelonggaran bea cukai, trik asmara, yang meliputi penjualan obat-obatan untuk meningkatkan kinerja seksual, kencan internet, petunjuk militer, iklan untuk pornografi, iklan untuk tujuan luar ruangan insidental, pekerjaan "telecommute" yang menjanjikan uang besar, terutama belanja online, perjanjian strategis, dan lainnya. Kemungkinan sorotan yang paling bernilai untuk penyaringan spam terdiri dari: Bingkai Pesan dan panjang pesan, pemeriksaan kemunculan kata, contoh pesan harian (email yang tidak diminta sebagian besar waktu memiliki banyak inkonsistensi semantik), usia penerima, jenis kelamin dan kebangsaan, penerima yang menanggapi (menunjukkan apakah penerima menjawab pesan), konten dewasa dan kumpulan frasa dari konten pesan.

Kemampuan akun pengirim yang diterapkan untuk isolasi email yang tidak diminta meliputi AS pengirim (perpindahan negara sebagaimana dinyatakan oleh pelanggan pada profil mereka dan sebagaimana diungkapkan melalui alamat IP mereka), alamat IP pengirim, email pengirim, usia pengirim dan penerima, popularitas pengirim. Sorotan yang kurang luas adalah pemisahan geografis antara pengirim dan penerima, tanggal lahir pengirim, nama pengguna dan frasa rahasia pengirim, harapan hidup akun, jenis kelamin pengirim, dan usia penerima. Pengakuan pesan email sampah dengan rentang sorotan paling sedikit cukup besar mengingat sifat komputasional dan waktu yang beragam. Pilihan sorotan terdiri dari dokumen seperti stemming, pengusiran keributan, dan langkah-langkah evakuasi kata berhenti.

Pengumpulan Spam Email yang Tersedia Secara Gratis

Kumpulan data yang terdapat dalam suatu kelompok mengasumsikan proses yang sangat besar dalam mensurvei pameran saluran spam apa pun. Terlepas dari kenyataan bahwa ada banyak kumpulan data adat yang umumnya diterapkan untuk mengelompokkan konten tekstual, beberapa ilmuwan dalam bidang penyaringan email sampah membuat pengumpulan yang diterapkan untuk menilai kecukupan saluran yang diusulkan yang dapat diakses oleh manusia dalam standar.

15.3 LANGKAH EVALUASI EKSEKUSI

Umumnya, jaringan Spam dievaluasi pada basis data besar yang berkaitan dengan email ham dan yang tidak diminta yang dapat diakses secara terbuka oleh klien. Ini adalah kasus kuantifikasi eksekusi yang dapat diterapkan adalah akurasi jenis (acc). Ini adalah variasi pesan yang relatif luas yang diurutkan dengan tepat, tingkat pesan yang dicirikan dengan baik digunakan sebagai ukuran tambahan untuk menilai eksekusi saluran. Namun, telah ditunjukkan bahwa menggunakan akurasi sebagai daftar eksekusi utama tidaklah cukup baik.

Pengukuran eksekusi yang berbeda, misalnya, ikhtisar, akurasi, dan ukuran tersirat yang digunakan di dalam bidang pemulihan informasi harus menjadi ide, jadi positif palsu dan negatif palsu juga diterapkan dalam hipotesis preferensi. Ini penting karena biaya yang terkait dengan kesalahan klasifikasi. Pada intinya, sementara pesan email yang tidak diminta salah diberi nama ham, ia menawarkan untuk naik ke beberapa derajat di samping kesulitan pokok, meskipun hal utama yang perlu dilakukan pelanggan adalah menghapus pesan tersebut. Sebaliknya, ketika pesan email yang tidak diminta ditandai dengan tidak tepat sebagai spam, ini menunjukkan bahaya hilangnya informasi yang cukup besar karena kesalahan pengelompokan saluran. Ini penting terutama dalam pengaturan di mana pesan email yang tidak diminta dihapus. Dengan cara ini, tidaklah cukup untuk menilai pameran mesin apa pun yang memperoleh pengetahuan tentang perhitungan yang diterapkan di saluran surat sampah yang memanfaatkan ketepatan pengelompokan secara menyeluruh. Selain itu, dalam pengaturan yang tidak proporsional atau sepihak di mana jumlah pesan spam yang digunakan untuk mencoba penyajian saluran jauh lebih baik daripada pesan ham, pengklasifikasi dapat melaporkan ketepatan yang tinggi melalui pemfokusan pada penemuan pesan surat sampah secara menyeluruh yang ditunjukkan pada Tabel 15.1.

Dalam situasi, di mana tidak selalu ada kemungkinan nol untuk salah mengklasifikasikan pesan "bukan spam" atau ham, merupakan suatu keharusan bahwa suatu perubahan harus dicapai di antara kedua jenis kesalahan, bergantung pada kecenderungan

konsumen dan penanda pameran yang diterapkan. Rumus untuk menghitung ketepatan kelas dan kesalahan kelas digambarkan dalam Persamaan (15.1) dan (15.2) di bawah :
Mari kita pertimbangkan:

N_h= Jumlah email bukan sampah yang akan diklasifikasikan.

N_s= Rentang email yang tidak diminta yang akan dikategorikan

$Klasifikasi\ Acc\ (Acc) = \|h \rightarrow h\| + \|s \rightarrow s\|$ **Acc = Ketepatan*	(15.1)

Tabel 15.1 Pengumpulan spam email yang tersedia secara bebas.

Nama Dataset	Spam	Non Spam	Tingkat Spam	Tahun
Spam simpanan	1590	0	100%	1998
Basis Spam	1813	2788	39%	1999
Ling-Spam	481	3412	17%	2000
PU1	481	618	44%	2000
PU2	1897	4150	31%	2001
PU3	142	579	20%	2002
ZH1	1826	2313	44%	2003
Gen-Spam	571	571	50%	2003
Trek 2005	1205	428	74%	2004
Orang Besar	31,196	9212	78%	2005
Korpus Phishing	52,790	39,399	57%	2005
Trek 2006	8549	0	100%	2005
Trek 2007	415	0	100%	2010

$$\text{Kesalahan Klasifikasi (Err)} = \frac{Nh + Ns}{Nh + Ns} = 1 - Acc = |h \rightarrow s| + |s \rightarrow h| \quad (15.2)$$

Akurasi dan Kesalahan Susunan biasanya mempertimbangkan keberadaan absolut terdistorsi $|h \rightarrow s|$ dan keberadaan minus terdistorsi $|s \rightarrow h|$ untuk menanggung biaya yang setara. Penting untuk diperhatikan bahwa biaya kesalahan yang tidak seimbang terlibat dalam penyaringan spam. Pengelompokan spam secara salah (atau disebut juga kejadian positif palsu) adalah kesalahan yang mahal dibandingkan dengan pesan spam yang hanya melewati saluran. Kejadian seperti itu disebut sebagai kejadian negatif palsu. Pada saat email asli diberi nama ham dengan benar, itu dikenal sebagai kejadian positif asli $|h \rightarrow h|$. Namun, ketika email spam diberi label yang tepat sebagai spam, maka kejadian negatif asli $|s \rightarrow s|$ telah terjadi. Mengingat klarifikasi di atas, *Distorted Absolute Rate* (DAR) dicirikan sebagai proporsi email

asli yang didelegasikan sebagai spam. Itu dimaksudkan dengan menggunakan persamaan dalam Persamaan (15.3) di bawah ini

$$\text{DAR} = \frac{\text{Jumlah Absolut Terdistorsi}}{\text{Jumlah Absolut Terdistorsi} + \text{Jumlah Negatif Benar}} \quad (15.3)$$

Selain itu, mengizinkan email yang tidak diminta yang telah disusupi malware, adware, spyware, Trojan, botnet, virus, worm, atau umpan phishing yang terdiri dari pesan yang mengaku berasal dari situs web sosial, situs web kencan, situs web lelang, bank, pemroses pembayaran daring biasanya digunakan untuk menjebak penerima. Hal ini membuat pelanggan rentan terhadap kerugian besar. Rasio pesan spam yang salah diberi label sebagai sah dikenal sebagai *Distorted Negative Rate* (DNR).

Itu adalah metrik yang sangat tepat untuk membandingkan kinerja filter. Rumus untuk menghitung yang adil ada dalam persamaan (15.4) di bawah ini, mengizinkan email spam yang telah disusupi malware termasuk pesan yang mengaku berasal dari situs web sosial, situs web kencan, situs web lelang, bank, pemroses pembayaran daring umumnya digunakan untuk menjebak korban. Rasio email yang tidak diminta yang salah diberi label sebagai sah disebut *Distorted-Negative Charge* (DNC). Itu adalah metrik yang sangat tepat untuk mengevaluasi kinerja keseluruhan filter. Sistem untuk menghitung DNR ada pada persamaan (15.4).

$$\text{DNR} = \frac{\text{Jumlah Negatif yang Terdistorsi}}{\text{Jumlah Absolut Sejati} + \text{Jumlah Negatif Terdistorsi}} \quad (15.4)$$

Filter spam dengan DAR dan DNR rendah memiliki kinerja yang baik. Karakteristik ini (DNR dan DAR) menunjukkan kinerja filter tanpa tujuan penundaan pada batas pemilihan klasifikasi tanpa menghasilkan estimasi probabilitas. Efisiensi filter yang memperkirakan peluang kondisional organisasi setelah melakukan klasifikasi didasarkan sepenuhnya pada kemungkinan yang diestimasikan yang dapat direpresentasikan melalui kurva yang dikenal sebagai Kurva ROC (Karakteristik Operasi Penerima). Kurva ROC adalah plot grafis yang menunjukkan fungsionalitas analitis filter email sampah saat tingkat biasnya berubah. Kurva ROC dihasilkan dengan menggunakan plot rasio positif nyata terhadap rasio positif palsu (DAR) pada satu pengaturan ambang batas. Rasio positif nyata disebut sebagai Sensitivitas. Rasio positif palsu disebut probabilitas alarm palsu yang dihitung dengan bantuan pengurangan rasio spesifisitas dari 1 (yaitu E. 1 - spesifisitas). Dua metrik yang diimpor dari sektor pemulihan statistik 'ingat' dan 'presisi' digunakan untuk memperoleh efektivitas dan fungsi filter spam. Karena alasan dimana:

$$|\text{S} \rightarrow \text{NS}| = \text{Counts of spam emails categorized as non} - \text{spam}$$
$$|\text{NS} \rightarrow \text{S}| = \text{Counts of non} - \text{spam emails named spam separately}$$

Dan di sini $|\text{Ns} \rightarrow \text{Ns}|$ dan $|\text{Ns} \rightarrow \text{S}|$ Persamaan (15.5) di bawah ini menggambarkan ‘ingatan’ spam (Rs) dan presisi spam (Ps):

$$R_S = \frac{|S \to S|}{|S \to S| + |S \to H|} \text{ and } P_S = \frac{|S \to S|}{|S \to S| + |H \to S|} \quad (15.5)$$

Ingat (Rs) yang juga dikenal sebagai efektivitas dapat didefinisikan sebagai variasi pesan spam yang relatif luas yang berhasil dicegah oleh filter agar tidak masuk ke kotak masuk. Selain itu, Presisi (Ps) dijelaskan karena keandalan filter dihitung dengan cara membagi rentang pesan yang diklasifikasikan dengan bantuan filter sebagai spam tetapi sebenarnya sah dengan menggunakan berbagai macam email.

Mengevaluasi kinerja filter email sampah yang berbeda, penggunaan (Rs) dan (Ps) adalah pemikiran sensitif tentang nilai luar biasa yang terlibat dalam perhitungan yang menghasilkan (Rs) dan (Ps). Nilai positif palsu adalah (lambda λ kali) daripada nilai negatif, di mana (λ) adalah komponen, yang bersifat numerik, yang menentukan seberapa 'mudah berubahnya' untuk mengklasifikasikan email yang sah sebagai email sampah. Pemahaman remunerasi harus diperhitungkan yang dapat dilakukan dengan menjadikan setiap email yang sah sama dengan α email. Dalam likuidasi ini, komponen yang digunakan untuk menghitung ukuran terkait harga yang terdiri dari Proporsi Biaya Lengkap (CCP) dan Akurasi Tertimbang (WAcc) dibahas.

Proporsi biaya lengkap digunakan untuk mengukur akurasi filter. CCP yang lebih tinggi menunjukkan kinerja keseluruhan yang lebih baik. Sementara biaya CCP < 1, jauh lebih baik untuk tidak menggunakan pembersihan. Dalam situasi di mana harga seimbang dengan waktu yang terbuang, CCP mengukur jumlah waktu yang terbuang oleh pengguna untuk menghapus semua email yang tidak diminta terlepas dari kenyataan bahwa filter email sampah telah disiapkan. Kemudian membandingkannya dengan waktu yang dihabiskan untuk membuang email yang tidak diminta secara manual yang terhindar dari filter selain waktu yang diperlukan untuk memulihkan dari pesan yang sah yang telah diblokir secara keliru. Dua kekuatan utama CCP adalah bahwa itu adalah dimensi yang sejauh ini tidak menikah, sementara sebagian besar ukuran sensitif biaya alternatif memerlukan setidaknya dua angka. Meskipun demikian, hal ini dapat memberikan efek yang salah tentang efektivitas filter karena CCP yang lebih tinggi dapat menunjukkan tingkat yang menurun secara signifikan atau tingkat yang sangat tinggi.

Demikian pula, CCP rentan terhadap stabilisasi pengumpulan. Stabilitas pengumpulan adalah situasi di mana tingkat surat yang tidak diminta dan tidak ada pesan spam di dalam rangkaian tersebut berbeda-beda. Portabilitas nilai-nilai tersebut merupakan salah satu kelemahan CCP. Selain itu, kontras dapat digambarkan dengan lebih baik di antara nilai-nilai CCP sementara semua CCP yang dievaluasi dihitung dengan cara λ yang sebanding. Rumus perhitungan untuk Akurasi Tertimbang (WAcc), Biaya Kesalahan Tertimbang (Wherr), dan Bagian Biaya Keseluruhan (CCP) direpresentasikan dalam Persamaan. (15.6), (15.7), dan (15.8) di bawah ini:

$$W_{Acc} = \frac{\aleph|H \to H| + |S \to S|}{N_H + N_S} \text{ and } W_{Acc} = 1 - W_{Acc} \quad (15.6.)$$

$$W_{Acc} = \frac{\aleph|H \to S| + |S \to H|}{N_H + N_S} \quad (15.7)$$

$$TCR = \frac{N_S}{\aleph|H \to S| + |S \to H} \quad (15.8)$$

Saat menghitung sensitivitas filter, lambda (λ) mengambil keputusan untuk salah mengklasifikasikan email non-spam sebagai email sampah. Nilai sensitivitas dengan komponen λ/(1 + λ) dimasukkan ke dalam brim. Versi tersebut dibangun kembali dan diukur pada berbagai gaya tingkat ketegasan λ. Ukuran akurasi pengujian dijelaskan sebagai harmonik tertimbang yang disarankan oleh presisi (Ps) dan memori (Rs) pengujian dalam suatu persamaan. F-measure menggunakan parameter yang memungkinkan negosiasi dicapai tentang recall dan presisi. F1 mewakili derajat f konvensional yang biasanya digunakan dan memberikan bobot yang seragam pada memori dan presisi seperti yang ditunjukkan pada Persamaan (15.9) dan (15.10).

$$F_1 = \frac{2 * \text{recall} * \text{precision}}{\text{Precision} + \text{Recall}} \tag{15.9}$$

$$F_{\text{Beta}} \frac{(1 + \text{Bet a}_2) * \text{recall} * \text{precision}}{\text{Recall} + \text{Bet a}_2 * \text{precision} + \text{Recall}} \tag{15.10}$$

Dalam keadaan di mana kita memiliki 0 < beta < 1, memberikan presisi yang lebih tinggi sementara, ketika kita memiliki beta > 1, memberikan signifikansi yang lebih besar. Secara dangkal, ukuran-f adalah kasus khusus dari derajat-f tertimbang ketika beta = 1.

15.4 KLASIFIKASI - TEKNIK PEMBELAJARAN MESIN UNTUK SPAM EMAIL

Kemudian, karakterisasi email yang tidak diminta pada dasarnya ditangani melalui kalkulasi AI (ML) yang disengaja untuk dibagi antara pesan email spam dan tidak diminta. Kalkulasi AI mencapai hal ini melalui pendekatan yang terprogram dan fleksibel. Berbeda dengan bergantung pada keputusan yang dikodekan secara manual yang tidak berdaya menghadapi sifat pesan email sampah yang terus berubah, strategi ML memiliki kapasitas untuk mendapatkan statistik dari berbagai pesan yang diberikan, dan kemudian, menggunakan statistik yang diperoleh untuk mengelompokkan pesan baru yang benar-benar diterimanya. Perhitungan ml memiliki kemampuan untuk melakukan lebih tinggi tergantung pada pengalaman mereka. Pada fase ini, kami akan mensurvei kemungkinan strategi AI paling terkenal yang telah dilakukan di area email sampah.

Teknik Flock - Pengelompokan

Pengelompokan adalah pengorganisasian sekumpulan pola ke dalam pelatihan terkait. Teknik pengelompokan digunakan dalam membagi investigasi kasus ke dalam kelompok yang cukup mirip yang disebut kluster. Strategi pengelompokan telah melibatkan banyak peneliti dan akademisi di berbagai bidang aplikasi. Algoritma pengelompokan yang dapat berupa perangkat pembelajaran tanpa pengawasan digunakan pada kumpulan data email sampah yang biasanya memiliki label asli.Dua jenis metode pengelompokan yang digunakan untuk email sampah adalah pengelompokan berbasis kepadatan dan *k-nearest neighbor* (KNN). Berbasis kepadatan adalah metode pengelompokan di mana pendekatan laporan lain telah dieksploitasi untuk membersihkan spam. Metode ini berpotensi untuk mengenkripsi pesan, yang pada gilirannya menegakkan kerahasiaan pesan.

KNN adalah metode berbasis distribusi, yang tidak lagi bergantung pada asumsi bahwa informasi diambil dari distribusi probabilitas tertentu. Hampir semua statistik terapan tidak

mematuhi postulat hipotetis yang biasa dibuat (yang mencakup kombinasi gaussian, dapat dipisahkan secara linier, dan lainnya). Algoritma non-parametrik seperti KNN dapat digunakan untuk menyelamatkan skenario semacam ini. Dalam pengklasifikasi KNN, model klasifikasi tidak selalu dibangun dari statistik, sebaliknya, kategori diselesaikan melalui pencocokan contoh uji dengan berbagai set data pelatihan, dan pilihan dibuat mengenai kelompok mana yang termasuk dengan mengandalkan kemiripan dengan rekan terdekat set data pelatihan. KNN juga dikenal sebagai pembelajar malas karena set data pelatihan tidak digunakan untuk melakukan generalisasi.

Berbagai set aturan KNN untuk memfilter email spam dijelaskan dalam algoritma di bawah ini. Di sini tetangga (r) mengembalikan k teman terdekat dari r, terdekat (r, t) kembali ke faktor terdekat dari t di r, dan lihat kelas (D) kembali ke label keanggunan s. Seperangkat aturan KNN yang mudah untuk jenis email yang tidak diminta ada di dalam sekumpulan aturan di bawah ini:

- Algoritma: Klasifikasi Email Spam
- Langkah 1. Pada bagian pertama algoritma ini temukan pesan email dengan label.
- Langkah 2. k adalah jumlah tetangga terdekat.
- Langkah 3. E menunjukkan Pesan Uji Email.
- Langkah 4. T menunjukkan sekumpulan Pesan Email Pelatihan.
- Langkah 5. Yang diberi label sebagai sekumpulan Pesan Email adalah L.
- Langkah 6. Tafsirkan berkas data pelatihan.
- Langkah 7. Tafsirkan berkas data pengujian.
- Langkah 8. untuk setiap r di E dan setiap t di T lakukan
- Langkah 9. Tetangga (r) = {}
- Langkah 10. Jika| Tetangga(r)| <k then
- Langkah 11. Tetangga (r) = Terdekat (r, t) U Tetangga(r)
- Langkah 12. Akhiri jika.
- Langkah 13. Jika| Tetangga(r)| >k then
- Langkah 14. Berisi (M,x_j,y_j)
- Langkah 15. akhiri jika.
- Langkah 16. akhiri untuk 17: kembalikan Klasifikasi Pesan Email Akhir (Spam/Email yang valid).
- Langkah 17. Akhiri.

Pengklasifikasi Naïve Bayes

Algoritma Klasifikasi Naïve Bayes dinamai menurut Thomas Bayes (1702–1761), yang membuat algoritma tersebut. Pengklasifikasi Bayesian menunjukkan teknik pembelajaran terbimbing dan teknik statistik untuk tipe. Ia bertindak sebagai model probabilistik yang memudahkan untuk mengungkap keraguan tentang model dengan cara yang etis melalui manipulasi probabilitas konsekuensi. Ia digunakan untuk menyediakan metode untuk masalah analitis dan prediktif.

Kategori tersebut memberikan pembelajaran yang masuk akal tentang algoritma dari keahlian sebelumnya dan fakta eksperimental yang dapat digabungkan. Tipe ini memberikan pandangan yang bermanfaat tentang pengetahuan dan membandingkan berbagai algoritma pembelajaran. Ia menghitung probabilitas yang tepat untuk prinsip dan kuat terhadap gangguan dalam fakta input. Pengklasifikasi Naïve Bayes adalah pengklasifikasi probabilistik

langsung yang didasarkan pada teorema Bayes dengan asumsi yang sifatnya netral. Ekspresi yang lebih baik untuk versi peluang adalah versi karakteristik mandiri yang ditunjukkan pada persamaan (15.11):

$$\text{Theorem: P (B given A)} = \text{P (A and B)/P (A)} \tag{15.11}$$

Konsep kedaulatan preventif kompleksitas dibuat untuk membuat perhitungan lebih sederhana, dan ini adalah ide penamaan algoritma sebagai 'naif'. Namun, rangkaian aturannya kuat dan sangat kokoh. Ia bekerja seperti algoritma terbimbing lainnya. Ada pertumbuhan dalam reputasi NB sebagai algoritma yang sederhana dan efisien secara komputasi dengan kinerja berkualitas tinggi dalam memecahkan masalah dunia nyata. Sebagai hasil dari kualitasnya yang luar biasa, pengklasifikasi NB telah menemukan perangkat lunak sebagai algoritma kategori dalam konten tekstual, email spam, evaluasi sentimen, sistem rekomendasi, opini spam, dan paket daring yang berbeda.

Pengklasifikasi Naif Bayes secara khusus digunakan dalam klasifikasi konten tekstual (karena menghasilkan hasil lanjutan dalam masalah multikelas dan aturan independensi) dan memiliki harga pemenuhan ekstra jika dibandingkan dengan beberapa algoritma pembelajaran gadget lainnya. Karena keuntungan yang jelas ini, mil-milnya diterapkan secara signifikan dalam subjek penyaringan surat yang tidak diminta (menemukan email surat yang tidak diminta) dan evaluasi sentimen (dalam evaluasi media sosial, untuk memahami evaluasi klien yang menguntungkan dan negatif). Penyaringan spam adalah penggunaan pengklasifikasi NB yang paling terkenal. Ini adalah metode yang populer untuk membedakan email surat yang tidak diminta dari email yang diautentikasi, yang disebut ham.

Sebagian besar pelanggan email menerapkan kalkulasi penyaringan spam Bayesian. Pada dasarnya semua prosedur pemisahan spam berbasis pengukuran menggunakan pengklasifikasi Naïve Bayes untuk mengelompokkan wawasan setiap token ke skor absolut dan skor tersebut digunakan dalam membuat tujuan pada penyaringan. Simbol T yang berarti peringkat spam diproses sebagaimana diuraikan dalam Persamaan (15.12)

$$S[T] = \frac{C_{spam}(T)}{C_{spam}(T) + C_{Ham}(T)} \tag{15.12}$$

Di mana

- $C_{Spam}(T)$ = Jumlah pesan spam yang berisi token T,
- $C_{Ham}(T)$ = Jumlah pesan ham yang berisi token T,

Ada kebutuhan untuk menggabungkan spamming token yang berbeda untuk menghitung spaminess pesan untuk menghitung kemungkinan pesan **m** dengan token **t1,..,tn**. Menghitung spaminess token tertentu dan membandingkannya dengan yang dibuat dari hammieness token tertentu adalah cara langsung untuk membuat klasifikasi. Hal ini direpresentasikan dalam persamaan (15.13) di bawah ini:

$$(H[M] = \prod_{i=1}^{N} (1 - S[T_I]) \tag{15.13}$$

Pesan tersebut didelegasikan sebagai spam jika item kedengkian total S [M] lebih besar daripada item kedengkian H[M]. Penggambaran di atas digunakan dalam kalkulasi pengelompokan Bayes untuk urutan email spam yang digambarkan di bawah ini:

Algoritma Kategori Naïve Bayes untuk Jenis Email yang Tidak Diminta

- Langkah 1. Masukkan kumpulan data email
- Langkah 2. Uraikan setiap email ke dalam token masalahnya.
- Langkah 3. Hitung peluang untuk setiap token S [w] = $C_{spam}(W)/(C_{ham}(W) + C_{spam}(W))$
- Langkah 4. Basis Data Menyimpan nilai kedengkian
- Langkah 5. untuk setiap pesan M, lakukan
- Langkah 6. selama (M tidak berakhir) lakukan
- Langkah 7. Periksa email untuk simbol Ti berikut.
- Langkah 8. Tanyakan basis data untuk kedengkian S(Ti).
- Langkah 9. Hitung kemungkinan email yang terkumpul S [M] dan H [M].
- Langkah 10. Hitung sinyal penyaringan email lengkap dengan: I [M] = f (S [M], H [M])
  Langkah 11. I[M]= I+S[M]−H[M]/2
- Langkah 12. if I [M] > threshold then
- Langkah 13. email ditandai sebagai spam
- Langkah 14. else
- Langkah 15. email ditandai sebagai non-spam.
- Langkah 16. end-if
- Langkah 17. end while
- Langkah 18. end for 19: return Final Email Classification (Spam/Valid email)
- Langkah 19. End.

Neural Network

Neural Network adalah kumpulan gadget penanganan tulus yang saling terhubung dan berkomunikasi satu sama lain melalui strategi untuk sejumlah besar asosiasi tertimbang. Setiap gadget mengakui kontribusi dari gadget tetangga dan sumber daya luar dan menghitung hasil yang dikomunikasikan ke kenalan yang berbeda. Mekanisme untuk mengkalibrasi massa asosiasi juga dibuat dapat diakses. Jaringan saraf adalah kalkulasi yang efektif untuk menangani masalah AI apa pun yang memerlukan karakterisasi. Karena kreativitasnya, mereka maju sebagai perangkat raksasa dalam pengaturan peralatan ilmuwan AI. Meskipun, struktur saraf biasanya tidak digunakan di dalam wilayah email sampah seperti yang mungkin dipikirkan orang.

Naïve Bayes adalah metode yang luar biasa untuk karakterisasi spam dengan presisi tinggi (99%) dan biaya denda palsu. Yang meningkatkan ketepatannya yang tinggi adalah sejumlah besar komponen persiapan yang saling berhubungan secara menyeluruh (neuron) yang dapat bekerja dalam penyelesaian untuk memberikan solusi bagi masalah tertentu. Misalnya, Google menggambarkan pertumbuhan presisi saluran spam Gmail dari 5% menjadi 99,9% setelah menggabungkan struktur saraf ke dalamnya. Ini menyimpulkan bahwa struktur saraf dapat berguna untuk meningkatkan kinerja saluran spam, terutama ketika dihibridisasi dengan urutan Bayesian dan metode unik. Di sisi lain, banyak ujian yang harus diselesaikan tentang penggunaan sistem saraf untuk penemuan surat yang tidak diminta, dan hampir semua penelitian pasang surut menganggap tata letak sistem, daya, dan kecepatan belajar menjadi konstan. Upaya eksplorasi tambahan perlu dipusatkan di sekitar kecukupan gadget di

seluruh kumpulan data daripada kepatutan berbagai rencana sistem untuk minat tersebut. Seperti yang ditunjukkan melalui, biasanya ada tiga jenis unit.

- Unit Input: menerima sinyal dari sumber luar.
- Unit Output: mengirimkan data dari luar ke komunitas.
- Unit Tersembunyi: menerima dan mengirimkan peringatan dalam jaringan

Operasi kerangka kerja disinkronkan sehingga sejumlah besar unit dapat bekerja secara setara. ANN diubah untuk mendapatkan banyak sumber informasi dan menghasilkan pengaturan hasil yang diperlukan. Siklus ini dikenal sebagai pembelajaran atau pelatihan. Ada dua jenis pelatihan dalam organisasi saraf (NN).

- Diawasi: organisasi dipersiapkan dengan memberikan banyak sumber informasi dan mengoordinasikan desain hasil, yang dikenal sebagai kumpulan data persiapan.
- Tidak diawasi: organisasi berlatih sendiri dengan mengumpulkan contoh.

Jaringan Syaraf Tiruan terdiri dari dua jenis yang biasanya tersirat setiap kali Jaringan Syaraf Tiruan digunakan. Yaitu persepsi perseptron dan persepsi berlapis. Segmen ini akan menjelaskan algoritma perseptron dan perangkat lunaknya untuk penyaringan surat sampah. Di bawahnya terdapat seperangkat aturan perseptron yang merupakan seperangkat aturan jaringan syaraf yang terkenal. Perseptron membantu menemukan fitur linier dari vektor karakteristik $f(x) = wTx + b$ sehingga $f(x) > 0$ untuk vektor satu organisasi, dan $f(x) < 0$ untuk vektor kelompok yang berbeda.

Selain itu, $w = (w1, w2, \ldots wm)$ adalah bobot karakteristik, di mana **b** adalah bias yang dimaksud. Kelompok-kelompok tersebut dapat diberi angka dari +1 dan -1, sehingga pencarian fungsi dilakukan. Pembelajaran perceptron dimulai dengan cara memilih parameter (w0, b0) secara acak dari resolusi $d(x) = \text{tanda}\,(wTx + b)$ dan memperbaruinya berulang kali. Set data pelatihan (x, c) dipilih pada pengulangan algoritma ke-n hingga volume yang karakteristik pilihan saat ini menggolongkannya sebagai salah (yaitu Tanda $(wnx + bn) \neq c$). Aturan praktis yang digambarkan oleh persamaan (15.14) di bawah ini digunakan dalam memperbarui parameter (wn, bn):

$$wn + 1 = wn + cx \qquad bn + 1 = bn + c \qquad (15.14)$$

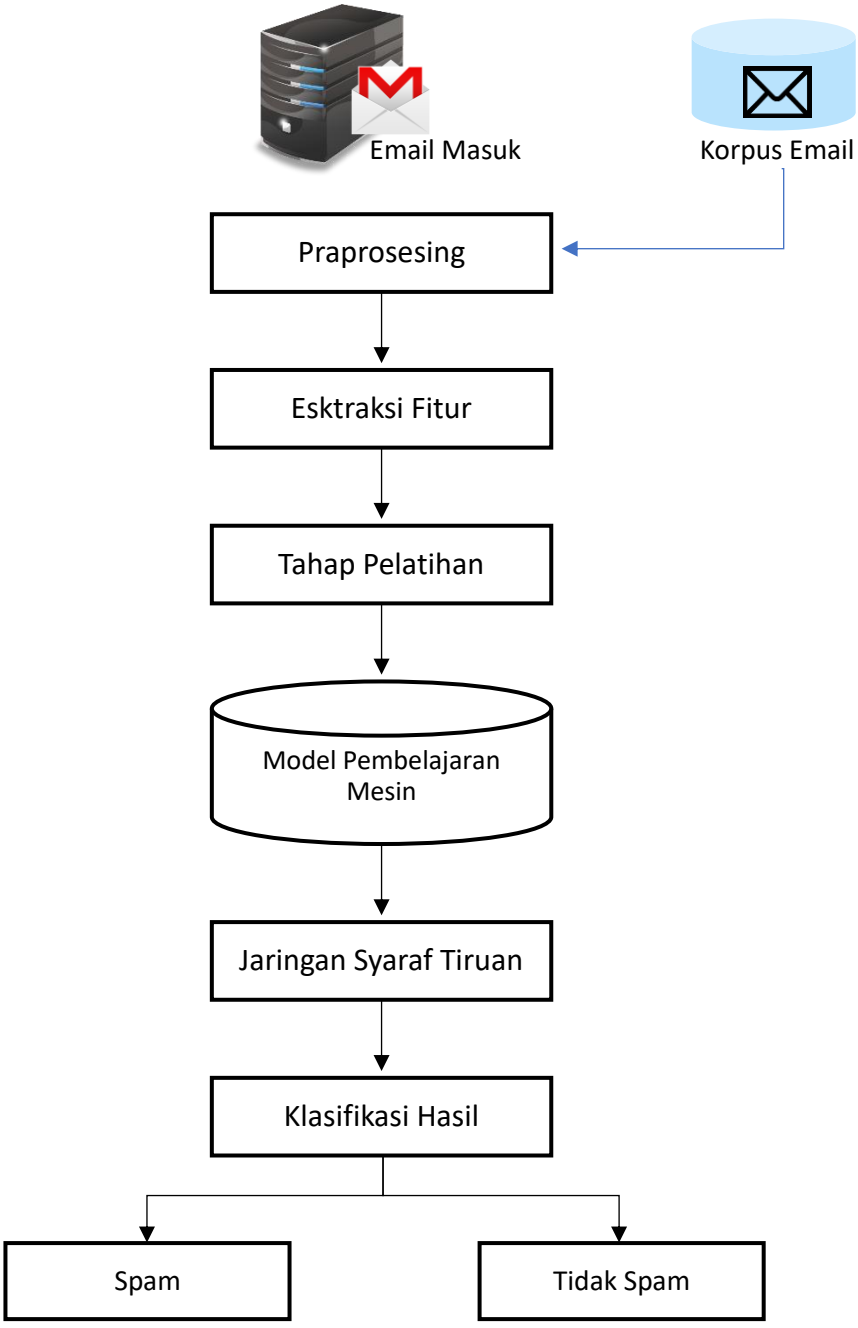
Standar untuk mengakhiri kalkulasi adalah bahwa kapasitas pilihan harus ditemukan yang secara tepat mengklasifikasikan semua uji persiapan ke dalam berbagai kelompok. Ada kalanya informasi persiapan tidak dapat diisolasi secara langsung, dalam kasus seperti itu langkah paling cerdik yang harus dilakukan adalah mengakhiri kalkulasi persiapan setelah jumlah informasi yang disusun secara tidak benar cukup sedikit. Algoritme di bawah ini berbicara tentang kalkulasi untuk Sistem Neural Perceptron untuk pengaturan spam email.

Perceptron Neural Community Set of Rules for Email Junk Mail Classification :

- Langkah 1: Masukkan contoh kumpulan data pesan email elektronik
- Langkah 2: Inisialisasi w dan b (ke nilai acak atau ke nol).
- Langkah 3: temukan contoh sekolah pesan (x,c) yang bertanda (wTx+ b).
- Langkah 4 jika tidak ada pola seperti itu, maka
- Langkah 5: pelatihan selesai
- Langkah 6: pertahankan w terakhir dan berhenti.
- Langkah 7: else

- Langkah 8: replace (w,b): w= w +cx, Langkah 9: b= b+ c
- Langkah 10: lanjut ke langkah delapan
- Langkah 11: end if
- Langkah 12: tentukan penetapan keanggunan pesan email elektronik (wTx+b)
- Langkah 13: kembali ke jenis pesan email terakhir (email sampah/email tidak diminta)
- Langkah 14: end.

Struktur pengklasifikasi email sampah Jaringan Syaraf digambarkan pada Gambar 15.3 di bawah ini.



Gambar 15.3 Struktur jaringan saraf (NN).

Algoritma Kunang-kunang

Kumpulan aturan kunang-kunang (FA) adalah kalkulasi heuristik met yang terutama berbasis populasi. Hal ini didasarkan pada perilaku berkilauan kunang-kunang. Kalkulasi tersebut mengacaukan dan meningkatkan beberapa pengaturan yang sedang naik daun dengan bantuan metode fisiognomi populasi untuk mengoordinasikan penyelidikan. Struktur kalkulasi tersebut ditetapkan pada penyelidikan gagasan korespondensi di antara kunang-kunang pada saat mereka bersiap untuk memiliki anggota keluarga seksual dan cepatnya mereka diberi kesempatan untuk mengambil risiko. Kunang-kunang berbagi informasi di antara mereka sendiri dengan menggunakan teknik untuk fitur berkilau mereka. Dengan sekitar 2.000 spesies kunang-kunang di bumi, semuanya menggunakan konfigurasi berkilau yang luar biasa. Kunang-kunang biasanya membuat sedikit kilauan dengan masalah perusahaan tertentu terhadap apa yang mereka hadapi.

Cahaya dihasilkan melalui masuknya cahaya secara biokimia oleh hewan hidup. Bergantung pada jenis cahaya, teman yang tepat akan memberikannya dengan meniru struktur yang sama atau membalas dengan menggunakan struktur yang asli. Sebaliknya, kekuatan cahaya berkurang karena jarak. Akibatnya, cahaya berkilauan yang terpancar dari kunang-kunang mendapat reaksi dari kunang-kunang di sekitarnya dalam lingkup cahaya yang terlihat. Tempat tinggal yang menarik dan pengembangan kunang-kunang mungkin ingin melewati perhitungan perampingan di mana persiapan mengikuti persiapan yang lebih baik (lebih berkualitas tinggi). Perhitungan firefly untuk pengaturan email yang tidak diminta ditetapkan sebagai berikut:

Algoritma 4. Set aturan firefly untuk klasifikasi email spam

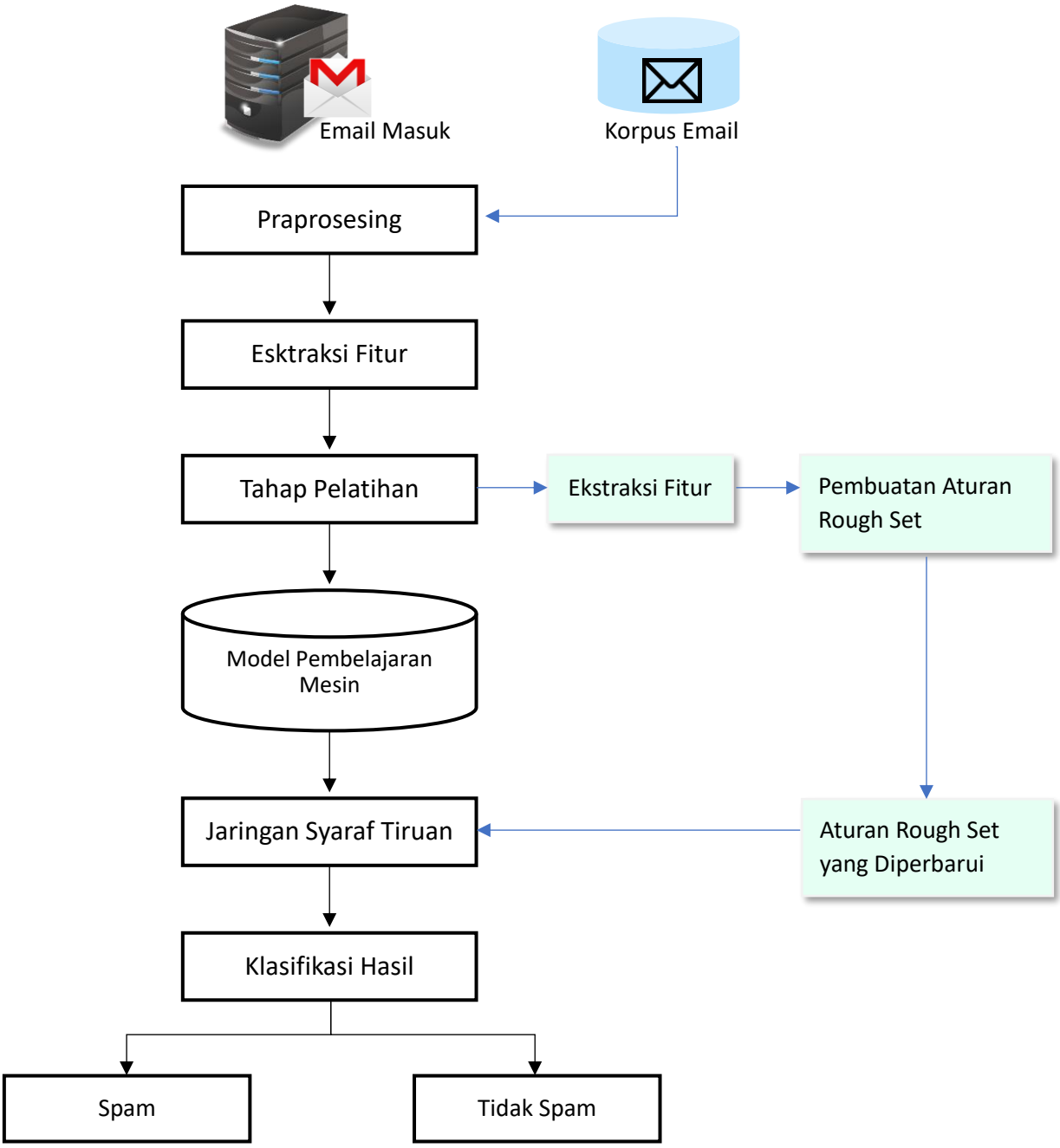
- Langkah 1. Masukkan rangkaian email dengan m variasi kemampuan yang luas.
- Langkah 2. Tetapkan nilai k sebagai 0, yaitu k = 0
- Langkah 3. Dapatkan populasi firefly n
- Langkah 4. Dapatkan jumlah atribut m.
- Langkah 5. Tetapkan populasi firefly
- Langkah 6. Untuk setiap firefly
- Langkah 7. Pilih firefly yang memiliki kesehatan yang baik
- Langkah 8. Pilih fitur yang setara dari bagian pemeriksaan koleksi email spam
- Langkah 9. Lihat pesan email.
- Langkah 10. k = k+1
- Langkah 11. Perbarui setiap firefly
- Langkah 12. Kategorikan pesan email sebagai spam dan email yang tidak diminta
- Langkah 13. akhiri untuk
- Langkah 14. Kembali ke kategori pesan email terakhir (email sampah/email bukan sampah) e-mail)
- Langkah 15. Akhiri

Pengklasifikasi Himpunan Fuzzy

Konsep himpunan fuzzy diusulkan pada tahun 1982 sebagai upaya untuk memperkenalkan struktur yang tepat untuk transformasi statistik menjadi pengetahuan secara robotik. Proses ini difokuskan pada pemecahan kategorisasi informasi yang diantisipasi, dipertanyakan, atau tidak lengkap yang diungkapkan sejauh data tersebut menjadi kenyataan. Konsep himpunan fuzzy dapat digambarkan sebagai strategi numerik yang berkelanjutan

untuk ketidakjelasan. Himpunan fuzzy sepenuhnya didasarkan pada konsep bahwa beberapa fakta dikaitkan dengan setiap item di alam semesta.

Rs adalah alat matematika yang mengkhususkan diri pada ketidakpastian. Hal ini sejalan dengan gagasan bahwa setiap versi yang salah dapat dievaluasi dari bawah dan dari atas dengan menggunakan asosiasi yang sifatnya halus. Salah satu tambahan besar dari konsep rs adalah kebutuhan untuk menemukan pengulangan dan kondisi di antara fitur-fitur. Konsep himpunan fuzzy telah diterapkan untuk menyaring email sampah karena menawarkan kalkulasi yang produktif dan tidak membosankan untuk mengekstraksi contoh-contoh tersembunyi dalam fakta. Ia juga memiliki kemampuan untuk menghubungkan berbagai strategi aktual reguler yang sulit untuk diurai.



Gambar 15.4 Alur kerja cara penyaringan surel himpunan fuzzy dari kotak surat perorangan.

Selain itu, ia mengakui penggunaan informasi kuantitatif dan subjektif. Ia memiliki kemampuan untuk memperkirakan kumpulan data minimum yang diinginkan untuk pekerjaan pengelompokan. Menemukan signifikansi statistik dan mengembangkan sekumpulan aturan keputusan dari kumpulan fakta yang diberikan adalah bagian dari kekuatan pengklasifikasi rs. Penting untuk diperhatikan bahwa ide himpunan yang lebat mengekspresikan ketidaktepatan melalui penggunaan fase garis batas dari suatu himpunan tetap dengan menggunakan cara keanggotaan.

Memiliki tempat marginal dari kekosongan tetap menyimpulkan bahwa himpunan tersebut telah dicirikan dengan jelas (aktual) jika himpunan tersebut tidak seharusnya kabur (tidak tepat). Untuk lokasi marginal yang terdiri dari setiap kejadian, satu faktor dalam himpunan tersebut menyiratkan bahwa apa yang kita anggap sebagai himpunan tidak cukup untuk menggambarkan himpunan tersebut secara tepat. Cenderung terlihat bahwa teori himpunan fuzzy memungkinkan klien untuk mengevaluasi keistimewaan catatan. Hal ini memungkinkan konsumen untuk membuat pengaturan persyaratan keinginan dari statistik secara jelas. Hal ini sederhana. Hal ini memberikan terjemahan langsung dari konsekuensi yang diberikan. Hal ini sesuai untuk pengelolaan simultan (sama/tersebarluaskan). Gambar 15.4 di bawah ini menunjukkan proses kerja pemisahan email dari teori himpunan fuzzy dari pelanggan yang menerbitkan wadah.

Support Vector Machine

Support Vector Machines (SVM) diawasi untuk memperoleh pengetahuan tentang algoritme yang berkinerja lebih tinggi daripada beberapa algoritme pembelajaran terkait lainnya. Aplikasi SVM dalam menyajikan solusi untuk masalah pemrograman kuadrat yang memiliki kendala ketidaksetaraan dan kesetaraan linier melalui pembedaan perusahaan-perusahaan yang unik melalui bidang hiper. SVM memanfaatkan sepenuhnya batas tersebut. Meskipun SVM mungkin tidak secepat teknik tipe lainnya, rangkaian aturan tersebut menarik kekuatannya dari akurasinya yang tinggi karena kemampuannya untuk memodelkan garis batas multidimensi yang tidak berurutan atau jujur. SVM bukannya tanpa masalah yang rentan terhadap skenario di mana suatu versi terlalu rumit beserta memiliki banyak parameter yang sebanding dengan rentang pengamatan. Kualitas-kualitas ini menjadikan SVM algoritme yang ideal untuk perangkat lunak di wilayah-wilayah seperti:

a) Pengenalan tulisan tangan virtual
b) Kategorisasi teks
c) Popularitas pembicara, dan sebagainya.

Sekarang, c menunjukkan parameter harga untuk mengatur kesalahan tampilan yang terjadi saat suatu fungsi secara hati-hati sesuai dengan kumpulan titik rekaman terbatas melalui penyempurnaan kesalahan ξ. Dalam perjalanan sekolah, kami berasumsi untuk memiliki seperangkat statistik untuk belajar, secara teoritis, mungkin hanya ada campuran parameter (c, γ) yang memiliki kapasitas untuk memberikan pengklasifikasi SVM yang unggul. Pada parameter, pencarian grid c dan γ adalah satu-satunya pendekatan yang mungkin ini diterapkan dalam pendidikan SVM untuk memperoleh penggabungan parameter. Estimasi rotasi k-fold aktif di dalam grid yang berusaha untuk memilih pengklasifikasi SVM dengan prediksi estimasi rotasi sempurna yang paling akurat. Algoritma pelatihan dan klasifikasi SVM untuk email spam ditawarkan dalam rangkaian aturan di bawah ini:

Algoritma 5.

- Langkah 1. Masukkan Contoh Pesan Email x untuk dipesan
- Langkah 2. Siapkan set S, sebuah karya, {c1, c2, … cn} dan {γ1, γ2, … γn}.
- Langkah 3. Jumlah tetangga terdekat k.
- Langkah 4. untuk I = 1 hingga n
- Langkah 5. tetapkan c=ci;
- Langkah 6. untuk j = 1 hingga q
- Langkah 7. tetapkan γ=γ;
- Langkah 8. Hasilkan pengklasifikasi SVM yang telah disiapkan f (x) yang mengaburkan batas penggabungan saat ini (C, γ);
- Langkah 9. jika (f (x) adalah karya diskriminan pertama yang dibuat) pada titik tersebut Langkah 10. pertahankan f (x) sebagai pengklasifikasi SVM terbaik f∗(x);
- Langkah 11. else
- Langkah 12. Bandingkan pengklasifikasi f (x) dan pengklasifikasi SVM terbaik saat ini f∗(x) dengan menggunakan persetujuan silang k-crease
- Langkah 13. pertahankan class1.
- Langkah 14. end if
- Langkah 15. end for
- Langkah 16. end for
- Langkah 17. Kembalikan Klasifikasi Pesan Email Akhir (Email Spam/Non-spam)
- Langkah 18. End

Pohon Keputusan

Pohon Keputusan (DK) adalah pengklasifikasi yang modelnya tampak seperti struktur pohon. Presentasi Pohon Pilihan adalah teknik yang tidak dapat disangkal yang mendorong perolehan data pada pengumpulan. Setiap hub dari DT adalah hub daun yang memutuskan penilaian segmen normal (kelas). Itu juga dapat menjadi hub pilihan yang menunjukkan tes tertentu yang akan dikoordinasikan pada penilaian suatu komponen, dengan satu cabang dan sub-pohon (yang merupakan bagian dari pohon yang lebih besar) yang menangani setiap kemungkinan hasil akhir dari tes tersebut.

Pohon pilihan dapat digunakan untuk menawarkan respons untuk masalah proses tindakan dengan memulai dari pembentukan pohon dan mengalaminya hingga mencapai hub daun yang memberikan hasil pengaturan. Pembelajaran Pohon Pilihan adalah filosofi yang telah diterapkan pada pemisahan spam. Faktanya adalah membuat model DT dan melatih model dengan tujuan untuk menebak penilaian variabel target yang difokuskan pada berbagai faktor data. Hub bagian dalam tertentu berbicara dengan bagian dari variabel informasi.

Daun tunggal menyiratkan penilaian variabel target yang diberikan bahwa penilaian faktor data berasal dari jalan yang mengarah dari akar ke daun. Dimungkinkan untuk merasa nyaman dengan pohon dengan memecah set utama menjadi beberapa subset berbeda tergantung pada penilaian komponen yang diberikan sebelumnya. Teknik ini diulang untuk setiap subset yang dihasilkan lebih dari sekali yang mengusulkan klarifikasi yang dikenal sebagai pembagian rekursif. Rekursi berhenti setelah semua subset pada hub tertentu semuanya memiliki faktor target yang sama. Standar lain yang dapat memicu akhir rekursi adalah saat mengisolasi set tidak semua juga memperbarui pengukur. Ada berbagai macam pohon keputusan seperti yang dijelaskan di bawah ini.

Pengklasifikasi NBTree

NBTree adalah jenis pohon keputusan yang menggabungkan Pohon Keputusan dengan pengklasifikasi Naïve Bayes di mana kekuatan masing-masing algoritma digabungkan. Metodologi ini bekerja dengan memanfaatkan Pengklasifikasi NBTree di node sementara pohon pilihan dibuat dengan satu variabel yang dipisahkan di setiap node. Untuk basis data yang berukuran besar, pengklasifikasi NBtree bermanfaat, jika ukuran basis data tidak seragam dan sorotannya tidak selalu dapat diatur sendiri, kualitas NBtree menjadi menonjol. Basis data email surat yang tidak diminta mengikuti contoh yang digambarkan di atas.

Algoritma Pohon Keputusan C4.5/J48

J48 adalah bentuk perhitungan pohon pilihan C4.5 yang disesuaikan, diurutkan ulang, dan dapat diakses secara terbuka. J48 dibuat dengan mempertimbangkan informasi di hub yang digunakan untuk menganalisis pentingnya karakteristik umum. Melalui pohon pilihan, model pohon dibuat dengan fuzzy penggunaan setiap komponen secara bergantian. Untuk mengerjakan ulang kumpulan data, perhitungan menggunakan estimasi elemen. Terlebih lagi, teruslah mencari wilayah dataset yang jelas memiliki satu kelas dan tampilkan wilayah tersebut sebagai daun.

Untuk sisa wilayah yang berisi kelas yang banyak, perhitungan memilih sorotan elektif. Perhitungan ini juga mempertahankan siklus pemisahan dengan hanya jumlah kejadian di wilayah tersebut yang akan datang saat daun sepenuhnya dibuat, atau tidak ada sorotan yang dapat digunakan untuk membuat setidaknya satu daun berfluktuasi di wilayah yang berbeda. Pohon pilihan yang dibuat oleh C4.5 dapat diterapkan untuk menangani masalah urutan yang berbeda. Perhitungan memilih sorotan yang juga dapat diisolasi menjadi subkelas di setiap hub. Hasil karakterisasi yang diperoleh ditandai dengan hub daun.

Induksi Pohon Versi Logistik (LVT)

LVT adalah pohon Keputusan yang menggunakan model regresi logistik pada simpul daun. Pengklasifikasi LVT telah membuktikan tingkat presisi dan kekokohan yang lebih baik di banyak wilayah instruksional. Versi logistik mudah didekodekan dibandingkan dengan pohon c4.5. Selain itu, telah dikonfirmasi bahwa kayu LVT dipadatkan ukurannya dibandingkan dengan pohon yang dibuat oleh c4.

Lima inisiasi. Tiga set aturan dikotomi iteratif (id3) yang terkenal yang diusulkan dengan bantuan Ross Quinlan dibahas untuk membangun pohon seleksi menggunakan entropi dan manfaat fakta. Entropi mengevaluasi pemalsuan serangkaian sampel email acak pada saat yang sama dengan manfaat informasi yang digunakan untuk menghitung entropi dengan cara membagi pola email dengan bantuan beberapa fungsi. Dengan asumsi kita memiliki kumpulan data email e dengan klasifikasi CJ, entropi telah menghitung penggunaan persamaan (15.15) di bawah ini.

$$\text{enttopy}\ (\mathrm{E}) \sum_{\mathrm{j}=1}^{|\mathrm{c}|} \Pr\ (\mathrm{c_j})\ \log_2 \Pr(\mathrm{C_J}) \tag{15.15}$$

Hubungan antara perolehan informasi dan entropi direpresentasikan dalam persamaan (15.16) di bawah.

$$\text{gain}\ (\mathrm{E}, \mathrm{F_i} = \text{entropy}\ (\mathrm{D}) - \text{entropy}_{\mathrm{Fi}}(\mathrm{E}) \tag{15.16}$$

Algoritma 6. Set aturan pohon keputusan untuk penyaringan email yang tidak diminta

- Langkah 1. Masukkan dataset pesan email
- Langkah 2. Hitung entropi untuk dataset
- Langkah 3. Pada saat yang sama seperti situasi lakukan
- Langkah 4. Untuk setiap atribut
- Langkah 5. Hitung entropi untuk semua nilai kategoris
- Langkah 6. Ambil entropi catatan rata-rata untuk karakteristik saat ini.
- Langkah 7. Hitung keuntungan untuk karakteristik saat ini
- Langkah 8. Pilih karakteristik manfaat terbaik
- Langkah 9. Menyerah untuk
- Langkah 10. Berhenti bahkan sebagai
- Langkah 11. Kembali ke kelas pesan email terakhir (email sampah/email bukan sampah)
- Langkah 12. Berhenti

Pengklasifikasi Ensemble

Pengklasifikasi Ensemble adalah pendekatan baru di mana serangkaian pengklasifikasi yang berbeda dilatih dan disusun untuk lebih meningkatkan akurasi kelas dari seluruh gadget pada kerumitan yang sama, dalam contoh ini untuk penyaringan email yang tidak diminta. Mereka adalah kategori algoritme pembelajaran gawai yang bekerja dalam penyelesaian dan dilakukan untuk meningkatkan kinerja kelas keseluruhan gawai. Pengklasifikasi ansambel yang paling umum adalah bagging dan boosting. Algoritme tersebut mengajarkan contoh pengklasifikasi pada berbagai subset dari keseluruhan set informasi. Bagging menggabungkan keluaran pengklasifikasi terampil pada sampel yang diambil dari pola set data yang lebih besar.

Boosting adalah pendekatan yang sangat ramah lingkungan yang menggabungkan rantai pendatang baru yang "lemah" untuk menciptakan pelajar tunggal yang lebih kuat daripada pelajar pria atau wanita. Ini dikategorikan sebagai pembelajaran serangkaian aturan yang berpusat pada gagasan hibridisasi berbagai hipotesis yang rentan, contoh yang sangat baik adalah perangkat AdaBoost. Tujuan boosting adalah untuk mendapatkan aturan kelas yang sepenuhnya benar melalui penggabungan beberapa aturan yang rentan atau hipotesis yang rentan yang masing-masing dapat menjadi sangat akurat.

Saat ini, boosting kini diterapkan di bidang kategori, regresi, pengenalan wajah, dan sebagainya. Algoritma penguat yang menggunakan proyeksi dengan tingkat keyakinan sedang diimplementasikan untuk memperbaiki masalah penyaringan surat yang tidak diminta. Karena kinerja keseluruhannya yang luar biasa dalam memperbaiki masalah klasifikasi, Adaboost merupakan perangkat yang umum digunakan untuk mempelajari seperangkat aturan. Adaboost cepat, seperangkat aturannya mudah dan bersih untuk diprogram, tidak adanya penyetelan parameter (kecuali t) membuatnya jauh lebih mudah.

Hutan Acak (RF)

Hutan acak adalah contoh teknik pembelajaran kolaboratif dan metode regresi yang cocok untuk memecahkan masalah yang berkaitan dengan pengklasifikasian catatan ke dalam perusahaan. Penggunaan pohon keputusan fuzzy pada algoritma ini mencakup estimasi. Pohon keputusan ini kemudian digunakan untuk usaha memprediksi lembaga; ini dilakukan

dengan mempertimbangkan kelompok-kelompok terpilih dari setiap pohon yang berbeda dan lembaga yang memiliki rentang suara terbaik diambil sebagai hasil akhirnya.

Metode ini telah mendapatkan pengakuan akhir-akhir ini dan telah menemukan perangkat lunak di bidang yang unik dan dalam literatur telah digunakan untuk menawarkan pendekatan terhadap masalah yang serupa. Kekuatan dari rangkaian aturan RF adalah biasanya memiliki lebih sedikit kesalahan penyortiran dan skor-f yang lebih besar dibandingkan dengan pohon pilihan. Meskipun faktanya bahwa hal itu jauh lebih mudah untuk diwujudkan bagi manusia, kinerja keseluruhannya umumnya sama atau bahkan lebih tinggi daripada SVM. Kinerjanya benar-benar luar biasa dengan unit statistik yang tidak merata yang dipertimbangkan melalui beberapa variabel yang hilang. Ia menyajikan mekanisme yang efisien untuk menghitung nilai perkiraan data yang hilang dan mempertahankan presisi dalam kondisi di mana persentase besar catatan hilang. RF memungkinkan konsumen untuk menumbuhkan pohon sebanyak mungkin.

Tingkat pelaksanaannya tinggi. Dalam banyak kasus di mana ukuran kumpulan informasi besar, banyak memori untuk penyimpanan fakta diperlukan. Menghitung kedekatan menunjukkan bahwa peningkatan dalam ruang penyimpanan yang dibutuhkan segera proporsional dengan jumlah kali yang dipercepat dengan bantuan jumlah semak. Proyek pengklasifikasian catatan baru dari vektor masukan dimulai dengan cara menempatkan vektor masukan di sepanjang setiap semak di hutan. Setiap pohon akan menjalankan kategorinya yang sering disebut sebagai pohon yang "memilih" lembaga tersebut. Hutan memilih lembaga mana yang memiliki suara terbanyak di dalam kawasan hutan.

Pada bagian pembahasan ini, kami mempelajari metode perolehan pengetahuan gadget dan perangkat lunaknya dalam konteks penyaringan surat sampah. Penilaian sejumlah algoritme telah diterapkan untuk kelas pesan sebagai surat sampah atau ham sebagaimana diberikan. Upaya yang dilakukan oleh para peneliti dalam memecahkan masalah pengklasifikasi email spam dengan pengetahuan gadget fuzzy disebutkan. Evolusi pesan surat sampah dari waktu ke waktu untuk menghindari filter juga diamati.

Tata letak struktural yang jelas dari filter surat sampah email dan taktik yang digunakan dalam menyaring email yang tidak diminta telah dicatat. Bab ini memetakan sejumlah kumpulan data dan metrik publik yang dapat digunakan untuk mengukur efektivitas pembersihan surat yang tidak diminta.

Tantangan algoritme pembelajaran mesin dalam menangani peluang surat sampah secara efisien telah ditunjukkan dan studi relatif dari teknik pembelajaran mesin yang tersedia dalam literatur telah diselesaikan. Selain itu, beberapa masalah studi terbuka yang terkait dengan filter surat sampah telah dieksplorasi. Literatur yang kami ulas menunjukkan bahwa kemajuan substansial telah dan masih terus terjadi dalam subjek tersebut. Karena masih banyaknya masalah dalam penyaringan surat sampah, penelitian lebih lanjut untuk mengevaluasi efektivitas penyaringan surat sampah perlu dilakukan. Hal ini dapat menjadikan peningkatan penyaringan surat sampah sebagai subjek penelitian penting bagi akademisi dan praktisi industri.

DAFTAR PUSTAKA

Alkattan, H., Subhi, A. A., Adelaja, O. A., Abotaleb, M., Mijwil, M. M., Mishra, P., ... & Turyasingura, B. (2023). Employing data mining techniques and machine learning models in classification of students' academic performance. *Babylonian Journal of Artificial Intelligence*, *2023*, 43-54.

Al-Shamiri, A. Y. R. (2021). Artificial intelligence and pattern recognition using data mining algorithms. *International Journal of Computer Science & Network Security*, *21*(7), 221-232.

Atzmueller, M., Fürnkranz, J., Kliegr, T., & Schmid, U. (2024). Explainable and interpretable machine learning and data mining. *Data Mining and Knowledge Discovery*, *38*(5), 2571-2595.

Balica, R. Ş., & Cuţitoi, A. C. (2022). Ethical Artificial Intelligence in Smart Mobility Technologies: Autonomous Driving Algorithms, Geospatial Data Mining Tools, and Ambient Sound Recognition Software. *Contemporary Readings in Law and Social Justice*, *14*(2), 64-81.

Binu, D., & Rajakumar, B. R. (Eds.). (2021). *Artificial intelligence in data mining: theories and applications*. Academic Press.

Chen, J., Yang, P., & Liang, Y. (2023, June). Big Data Mining Algorithm of Internet of Things based on artificial intelligence technology. In *2023 2nd International Conference on Artificial Intelligence and Blockchain Technology (AIBT)* (pp. 113-118). IEEE.

Cug, J., Trnka, M., & Popescu, G. H. (2023). Blockchain-based decentralized metaverse systems, industrial artificial intelligence of things, and spatial data mining and acoustic environment recognition algorithms in realistic 3D simulation environments. *Economics, Management and Financial Markets*, *18*(1), 57-72.

Deroy, A., Bailung, N. K., Ghosh, K., Ghosh, S., & Chakraborty, A. (2024). Artificial intelligence (ai) in legal data mining. *arXiv preprint arXiv:2405.14707*.

Dogan, M. E., Goru Dogan, T., & Bozkurt, A. (2023). The use of artificial intelligence (AI) in online learning and distance education processes: A systematic review of empirical studies. *Applied sciences*, *13*(5), 3056.

Doss, A. N., Maurya, N., Guru, K., Masood, G., Jaiswal, S., & Naved, M. (2022, June). The Impact of Data Mining and Artificial Intelligence on Supply Chain Management and Environmental Performance. In *Proceedings of Second International Conference in Mechanical and Energy Technology: ICMET 2021, India* (pp. 503-511). Singapore: Springer Nature Singapore.

Entezari, A., Aslani, A., Zahedi, R., & Noorollahi, Y. (2023). Artificial intelligence and machine learning in energy systems: A bibliographic perspective. *Energy Strategy Reviews*, *45*, 101017.

Gordan, M., Sabbagh-Yazdi, S. R., Ismail, Z., Ghaedi, K., Carroll, P., McCrum, D., & Samali, B. (2022). State-of-the-art review on advancements of data mining in structural health monitoring. *Measurement*, *193*, 110939.

Haue, A. D., Hjaltelin, J. X., Holm, P. C., & Placido, D. (2024). Artificial intelligence-aided data mining of medical records for cancer detection and screening. *The Lancet Oncology*, *25*(12), e694-e703.

Himeur, Y., Rimal, B., Tiwary, A., & Amira, A. (2022). Using artificial intelligence and data fusion for environmental monitoring: A review and future perspectives. *Information Fusion*, *86*, 44-75.

Khan, B., Hasan, A., Pandey, D., Ventayen, R. J. M., Pandey, B. K., & Gowwrii, G. (2021). Fusion of datamining and artificial intelligence in prediction of hazardous road accidents. In *Machine learning and iot for intelligent systems and smart applications* (pp. 201-223). CRC Press.

Li, J., Herdem, M. S., Nathwani, J., & Wen, J. Z. (2023). Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management. *Energy and AI*, *11*, 100208.

Ma, Y., Zhu, H., Yang, Z., & Wang, D. (2022). Optimizing the prognostic model of cervical cancer based on artificial intelligence algorithm and data mining technology. *Wireless Communications and Mobile Computing*, *2022*(1), 5908686.

Mayet, A. M., Salama, A. S., Alizadeh, S. M., Nesic, S., Guerrero, J. W. G., Eftekhari-Zadeh, E., ... & Iliyasu, A. M. (2022). Applying data mining and artificial intelligence techniques for high precision measuring of the two-phase flow's characteristics independent of the pipe's scale layer. *Electronics*, *11*(3), 459.

Menaga, D., & Saravanan, S. (2021). Application of artificial intelligence in the perspective of data mining. In *Artificial Intelligence in Data Mining* (pp. 133-154). Academic Press.

Paydar, S., Parva, E., Ghahramani, Z., Pourahmad, S., Shayan, L., Mohammadkarimi, V., & Sabetian, G. (2021). Do clinical and paraclinical findings have the power to predict critical conditions of injured patients after traumatic injury resuscitation? Using data mining artificial intelligence. *Chinese Journal of Traumatology*, *24*(01), 48-52.

Rashid, M., Strakova, J., & Valaskova, K. (2023). Geolocation Data Mining and Tracking, Generative Artificial Intelligence and Haptic and Biometric Sensor Technologies, and Network Visual and Employee Engagement Analytics in 3D Immersive Spaces. *Contemporary Readings in Law & Social Justice*, *15*(2).

Ruifeng, S. (2021). Research on data mining system based on artificial intelligence and improved genetic algorithm. *Journal of Intelligent & Fuzzy Systems*, *40*(4), 6731-6742.

Saleh, A. I., & Rabie, A. H. (2023). Human monkeypox diagnose (HMD) strategy based on data mining and artificial intelligence techniques. *Computers in Biology and Medicine*, *152*, 106383.

Shen, L. (2021). Data mining artificial intelligence technology for college English test framework and performance analysis system. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 3489-3499.

Simionescu, C., Danubianu, M., & Maciuca, M. S. (2023). How Data Mining and Artificial Intelligence Can Contribute to Increasing Academic Performance. *Didactica Danubiensis*, *3*(1), 72-85.

Sohail, A. (2023). Genetic algorithms in the fields of artificial intelligence and data sciences. *Annals of Data Science*, *10*(4), 1007-1018.

Strowel, A., & Ducato, R. (2021). Artificial intelligence and text and data mining: a copyright carol. In *The Routledge handbook of EU copyright law* (pp. 299-316). Routledge.

Torrentira Jr, M. C. (2024). Capabilities and application of artificial intelligence (AI) models in qualitative and quantitative data mining, data processing and data analysis. *European Journal of Education Studies*, *11*(9).

Ullrich, A., Vladova, G., Eigelshoven, F., & Renz, A. (2022). Data mining of scientific research on artificial intelligence in teaching and administration in higher education institutions: a bibliometrics analysis and recommendation for future research. *Discover Artificial Intelligence*, *2*(1), 16.

Ye, Z., & Su, L. (2021). The use of data mining and artificial intelligence technology in art colors and graph and images of computer vision under 6G internet of things communication. *International Journal of System Assurance Engineering and Management*, *12*(4), 689-695.

Zhang, P., Zheng, J., Lin, H., Liu, C., Zhao, Z., & Li, C. (2023). Vehicle trajectory data mining for artificial intelligence and real-time traffic information extraction. *IEEE Transactions on Intelligent Transportation Systems*, *24*(11), 13088-13098.

Zhang, S., & Duan, C. (2022). Clustering optimization algorithm for data mining based on artificial intelligence neural network. *Wireless Communications and Mobile Computing*, *2022*(1), 1304951.

Zia, A., Aziz, M., Popa, I., Khan, S. A., Hamedani, A. F., & Asif, A. R. (2022). Artificial intelligence-based medical data mining. *Journal of Personalized Medicine*, *12*(9), 1359.

ARTIFICIAL INTELLIGENCE DAN DATA MINING

Dalam Kerangka Sekuriti

Dr. Joseph Teguh Santoso, S.Kom, M.Kom.

BIODATA PENULIS

Dr. Joseph Teguh Santoso, M.Kom memiliki Jabatan Akademik Lektor Kepala dan praktisi industri yang berpengalaman. Saat ini menjabat sebagai Rektor Universitas Sains dan Teknologi Komputer (Universitas STEKOM), salah satu universitas terkemuka di Jawa Tengah, Indonesia. Dengan pengalaman lebih dari 13 tahun di dunia bisnis dan praktisi industri di China, beliau membawa perspektif global dan inovasi yang signifikan ke dalam dunia akademis.Sebagai seorang entrepreneur, penulis adalah pencipta TopLoker.com, sebuah platform inovatif yang merevolusi cara mencari dan menawarkan pekerjaan. TopLoker.com adalah portal lowongan bursa kerja terbesar di Indonesia, khusus untuk pendidikan SMA/SMK sederajat. TopLoker.com telah mendapatkan penghargaan sebagai juara 1 Startup4industry 2022 oleh Kementerian Perindustrian Republik Indonesia. Kontribusi Dr. Joseph dalam menyediakan akses pekerjaan yang luas bagi lulusan SMA/SMK telah membantu banyak individu menemukan peluang kerja yang sesuai dengan keahlian mereka. Selain itu, Dr. Joseph Teguh Santoso, M.Kom adalah pendiri dari dua organisasi yaitu (1) organisasi guru/pendidik PTIC (Perkumpulan Teacherpreneur Indonesia Cerdas) yang bertujuan untuk meningkatkan kualitas pendidikan dan kesejahteraan guru/pendidik dengan wawasan entrepreneurship, serta (2) organisasi industri PERKIVI (Perkumpulan Komunitas Industri dan Vokasi Indonesia) yang berfokus pada pengembangan link and match antara industri dan dunia pendidikan. Sebagai Rektor, Dr. Joseph Teguh Santoso, M.Kom memiliki kepemimpinan yang berorientasi pada hasil, dan berkomitmen untuk mendorong kemajuan Universitas Sains dan Teknologi Komputer (Universitas STEKOM). Saat ini Universitas STEKOM telah mengalami transformasi positif dalam peningkatan kualitas pendidikan, perluasan fasilitas, serta penguatan kemitraan Perguruan Tinggi Nasional dan Internasional. Beliau memprioritaskan pengembangan sumber daya manusia dan penelitian, serta memastikan bahwa universitas berada di garis depan dalam inovasi dan teknologi untuk mencapai tujuan akhir, yaitu lulusan yang mampu bekerja dan sukses setelah lulus. Dr. Joseph Teguh Santoso, M.Kom sering diundang sebagai pembicara di berbagai konferensi nasional maupun internasional dan telah menerima berbagai penghargaan atas dedikasinya dalam bidang pendidikan, industri, dan kewirausahaan.

YAYASAN PRIMA AGUS TEKNIK

PENERBIT :
YAYASAN PRIMA AGUS TEKNIK
Jl. Majapahit No. 605 Semarang
Telp. (024) 6723456. Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

ISBN 978-634-7227-26-3 (PDF)
9 786347 227263