BAB II

LANDASAN TEORI

2.1 Definisi Evaluasi

Menurut Arikunto (2009) evaluasi adalah kegiatan untuk mengumpulkan informasi tentang bekerjanya sesuatu, yang selanjutnya informasi tersebut digunakan untuk menentukan alternatif yang tepat dalam mengambil keputusan, sedangkan menurut Wakhinuddin (2009) dari aspek pelaksanaan, evaluasi adalah keseluruhan kegiatan pengumpulan data dan informasi, pengolahan, penafsiran, dan pertimbangan untuk membuat keputusan. Evaluasi adalah penilaian secara sistematis, mencakup pemberian nilai, atribut, apresiasi, pengenalan masalah, dan pemberian solusi atas permasalahan yang ditemukan.

Pengertian-pengertian di atas dapat disimpulkan bahwa evaluasi adalah serangkaian proses pengumpulan dan penilaian informasi terencana, untuk menilai suatu permasalahan yang terjadi dan hasilnya akan berguna dalam menentukan alternatif pengambilan keputusan untuk memberi solusi dari permasalahan yang dinilai.

2.2 Definisi Sistem

Gondodiyoto (2007) menyatakan bahwa sistem merupakan suatu kesatuan yang terdiri dari komponen-komponen atau subsistem yang berorientasi untuk mencapai tujuan tertentu.

Mc Leod (2004) mengemukakan *"A system is a group of elements that are integrated with the common purpose of achieving an objevtive"*. Secara garis besar dapat diartikan bahwa sistem adalah sekelompok elemen yang terintegrasi dengan maksud yang sama untuk mencapai suatu tujuan.

Jogiyanto (2009) berpendapat sistem adalah suatu jaringan kerja prosedur-prosedur yang saling berhubungan, berkumpul bersama-sama untuk melakukan suatu kegiatan atau untuk menyelesaikan suatu sasaran tertentu.

Maksud dari rangkaian kata prosedur pada definisi di atas adalah

rangkaian kegiatan input, proses, dan output yang saling bekerja sama untuk mencapai suatu tujuan tertentu. Input merupakan bahan baku yang dimasukkan ke dalam evaluasi akan menghasilkan gambaran kinerja sistem akademik persemesternya, khususnya untuk mahasiswa.

Tujuannya adalah untuk menentukan nilai akademik mahasiswa, diharapkan sistem, proses adalah suatu sistem yang mempunyai suatu pengolah yang akan merubah input menjadi keluaran atau output, sedangkan output adalah hasil keluaran dari energi yang telah diolah yang dapat dipergunakan oleh pihak lain dan diklasifikasikan menjadi keluaran yang berguna. Data yang digunakan untuk menghasilkan informasi tersebut membutuhkan media untuk penyimpanan yang dikenal dengan sebutan database.

2.3 Sistem Evaluasi

Sistem evaluasi adalah suatu sistem yang melakukan kegiatan untuk melakukan penilaian pada seluruh kegiatan akademik yang telah berjalan. Hasil dapat menjadi peringatan dan motivasi pada mahasiswa yang mempunyai nilai yang bermasalah untuk lebih giat dan serius. Memberikan informasi kepada mahasiswa, Kaprodi, dan Dosen Wali, tentang pengelompokan penilaian akademik mahasiswa bermasalah.

2.4 Data Mining

2.4.1 Definisi Data Mining

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data mining adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database yang besar. Data mining sering disebut juga *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola, atau hubungan dalam data set berukuran besar. Keluaran data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan.

Kehadiran data mining dilatarbelakangi oleh berlimpahnya data

yang dialami oleh berbagai institusi, perusahaan atau organisasi. Berlimpahnya data ini merupakan akumulasi data transaksi yang terekam beberapa waktu lamanya. Data mining juga dilatarbelakangi oleh atau adanya ledakan informasi dari berbagai media terutama internet. Data–data tersebut merupakan data transaksi yang umumnya diproses menggunakan aplikasi komputer. Pertumbuhan yang pesat dari akumulasi data atau informasi itu telah menciptakan kondisi suatu institusi memiliki bergunung-gunung data tetapi miskin informasi yang bermaanfaat, dan data mining hadir untuk menjawab tantangan tersebut.

2.4.2 Tahap-tahap Data Mining

Tan (2006) mengemukakan data mining sesungguhnya merupakan salah satu rangkaian dari proses pencarian pengetahuan pada database KDD. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interprestasi dan visualisasi dari pola-pola sejumlah kumpulan data. KDD adalah keseluruhan proses untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, pola yang ditemukan bersifat sah, baru, bermanfaat dan dapat dimengerti. Serangkaian proses tersebut yang memiliki tahap sebagai berikut :

1. Pembersihan data dan integrasi data (*cleaning and integration*)
   Proses ini digunakan untuk membuang data yang tidak konsisten dan bersifat noise dari data yang terdapat di berbagai database yang mungkin berbeda format maupun platform yang kemudian diintegrasikan dalam satu database data warehouse. Data warehouse merupakan tempat penyimpanan informasi yang dikumpulkan dari berbagai sumber, disimpan dalam skema yang dipersatukan (*unified schema*) dan biasanya bertempat pada tempat penyimpanan tunggal.
2. Seleksi dan transformasi data (*selection and transformation*)
   Data yang terdapat dalam database data warehouse kemudian direduksi dengan berbagai teknik. Proses reduksi diperlukan untuk mendapatkan hasil yang lebih akurat dan mengurangi waktu komputasi terutama untuk masalah dengan skala besar. Beberapa

cara seleksi, antra lain:

a. *Sampling*, adalah seleksi subset representatif dari populasi data yang besar.
b. *Denoising*, adalah proses menghilangkan noise dari data yang akan ditransformasikan.
c. *Feature extraction*, adalah proses membuka spesifikasi data yang signifikan dalam konteks tertentu.

Transformasi data diperlukan sebagai tahap pre-processing, dimana data yang diolah siap untuk ditambang. Beberapa cara transformasi, antara lain:

a. *Centering*, mengurangi setiap data dengan rata-rata dari setiap atribut yang ada.
b. *Normalisation*, membagi setiap data yang di centering dengan standar deviasi dari atribut bersangkutan.
c. *Scaling*, mengubah data sehingga berada dalam skala tertentu.

2.5 Klasifikasi

Han & Kamber (2006) mendefinisikan klasifikasi dan prediksi adalah dua bentuk analisis data yang bisa digunakan untuk mengekstrak model dari data yang berisi kelas-kelas atau untuk memprediksi trend data yang akan datang. Klasifikasi memprediksi data dalam bentuk kategori, sedangkan prediksi memodelkan fungsi-fungsi dari nilai yang kontinyu. Misalnya model klasifikasi bisa dibuat untuk mengelompokkan aplikasi peminjaman pada bank apakah berisiko atau aman, sedangkan model prediksi bisa dibuat untuk memprediksi pengeluaran untuk membeli peralatan komputer dari pelanggan potensial berdasarkan pendapatan dan lokasi tinggalnya. Prediksi bisa dipandang sebagai pembentukan dan penggunaan model untuk menguji kelas dari sampel yang tidak berlabel, atau menguji nilai atau rentang nilai dari suatu atribut. Dalam pendangan ini, klasifikasi dan regresi adalah dua jenis masalah prediksi, dimana klasifikasi digunakan untuk memprediksi nilai-nilai diskrit atau nominal, sedangkan regresi digunakan untuk memprediksi nilai-nilai yang kontinyu. Untuk selanjutnya penggunaan istilah *prediction* untuk

memprediksi kelas yang berlabel disebut *classification*, dan penggunaan istilah prediksi untuk memprediksi nilai-nilai yang kontinyu sebagai *prediction*.

Data input untuk klasifikasi adalah koleksi dari record. Setiap record dikenal sebagai instance atau contoh, ditentukan oleh sebuah tuple (x,y), dimana x adalah himpunan atribut dan y adalah atribut tertentu, yang dinyatakan sebagai label kelas (juga dikenal sebagai kategori atau atribut target). Klasifikasi adalah tugas pembelajaran sebuah fungsi target f yang memetakan setiap himpunan atribut x ke salah satu label kelas y yang telah didefinisikan sebelumnya. Fungsi target juga dikenal secara informal sebagai model klasifikasi.

Model klasifikasi berguna untuk keperluan berikut :

Pemodelan Deskriptif. Model klasifikasi dapat bertindak sebagai alat penjelas untuk membedakan objek-objek dari kelas-kelas yang berbeda. Sebagai contoh untuk para ahli Biologi, model deskriptif yang meringkas data.

Pemodelan Prediktif. Model klasifikasi juga dapat digunakan untuk memprediksi label kelas dari record yang tidak diketahui. Model klasifikasi dapat dipandang sebagai kotak hitam yang secara otomatis memberikan sebuah label ketika dipresentasikan dengan himpunan atribut dari record yang tidak diketahui.

2.6 Anomali Data

Data set terdiri dari sejumlah data yang setiap data mempunyai sejumlah fitur yang mendeskripsikan karakter data tersebut. Dalam data mining data set tersebut dapat diolah untuk menghasilkan informasi yang berguna, seperti klasifikasi, clustering, dan sebagainya. Tidak jarang dalam data-data yang akan diolah ditemukan adanya data yang karakteristiknya secara signifikan menyimpang/berbeda dengan karakteristik data pada umumnya.

Menurut Prasetyo (2012) data-data yang mempunyai karakter seperti ini sering disebut sebagai penyimpangan (*anomaly/outlier/noise*). Jika data yang menyimpang disebut sebagai outlier, maka data yang berada dalam daerah yang wajar/normal maka disebut sebagai inlier. Pada kasus clustering,

kehadiran outlier dapat memberikan hasil clustering tidak maksimal. Umumnya penghilangan outlier (*outlier removal*) menjadi pemrosesan awal (*pre-processing*) pada data set agar memberikan hasil yang baik.

Sedangkan pada kasus-kasus tertentu berguna untuk mendeteksi keadaan yang tidak biasa pada data yang didapat. Pada bab ini, istilah anomali data, outlier, dan noise yang digunakan mempunyai maksud yang sama.

Outlier biasanya dianggap sebagai obyek/data yang jumlahnya sangat kecil jika dibandingkan dengan data normal lainnya, misalnya probabilitas kemunculannya satu dari seribu data, tetapi bisa menjadi seribu jika data sudah berjumlah satu juta. Dengan demikian, pekerjaan deteksi outlier pada data yang menyimpang menjadi pekerjaan yang penting untuk berbagai keperluan dalam data mining.

Dalam dunia nyata, penyimpangan data dapat dilihat pada sebuah contoh kasus penyimpanan data usia manusia, umumnya usia manusia mulai 0 sampai 90 tahun, hal ini tidak menutup kemungkinan bahwa ada juga manusia yang berusia diatas 90 tahun, misalnya 100 tahun atau 110 tahun. Data manusia yang usianya menyimpang dari data usia pada umumnya disebut anomali/outlier/noise. Data seperti ini kadang disebut juga pengecualian data.

2.6.1 Penerapan Teknik Anomali Data

Penerapan deteksi anomali, dapat ditemukan pada sejumlah bidang-bidang kegunaan seperti: deteksi penggelapan (*fraud detection)*, deteksi penyusupan (*intrusion detection),* gangguan ekosistem (*ecosystem disturbances*), kesehatan masyarakat (*public health*), dan kedokteran. Berikut penjelasan penggunaan deteksi anomali dalam bidang tersebut sebagai berikut :

a. Deteksi penggelapan (*fraud detection*)

   Pembelian barang secara online dengan kartu kredit memicu munculnya para pencuri kartu kredit yang menggunakan nomor kartu kredit pelanggannya dengan membeli barang-barang yang biasanya mempunyai jenis, jumlah, pola pembelian yang berbeda dengan pelanggan yang seharusnya. Tanpa ada usaha dari bank penyedia layanan kartu kredit untuk secara berkala

melakukan deteksi penyimpangan pola transaksi kartu kredit setiap pelanggannya tentu penggelapan isi kartu kredit yang dilakukan pencuri akan merugikan pelanggannya.

b. Deteksi penyusupan (*intrusion detection*)

Akses ke sumber data dalam instansi, baik komputer maupun jaringan tidak dapat dibantah bahwa datangnya dari tempat umum (internet), tetapi untuk akses yang bisanya bertujuan misalnya untuk mematikan fungsi server atau merusak data yang tersimpan akan memiliki perilaku yang berbeda, misalnya dari isi header protokol atau isi pesan tertentu yang disisipkan didalamnya. Dengan mengintegrasikan metode deteksi anomali pada protokol jaringan seperti firewall atau router, maka diharapkan dapat meningkatkan sistem keamanan jaringan.

c. Gangguan ekosistem (*ecosystem disturbances*)

Dalam kehidupan di bumi, ada sejumlah pola-pola kehidupan seperti musim, pola kehidupan hewan, yang biasanya ada pola tertentu yang mengalami penyimpangan. Penyimpangan ini dapat mengakibatkan adanya gejala alam yang tidak biasa terjadi. Misalnya, pada masalah penyimpangan pola musim yang dapat mengakibatkan pola kehidupan hewan tertentu menjadi berubah. Deteksi penyimpangan kondisi alam sangat penting untuk dilakukan agar dapat mengetahui sejak dini bahaya-bahaya yang mungkin bisa terjadi, seperti : banjir, kebakaran, gempa, pencemaran, dan sebagainya.

d. Kesehatan masyarakat (*public health*)

Di Indonesia, ada puskesmas di setiap kecamatan atau kabupaten, data-data rekam medis yang tersimpan di setiap puskesmas mencatat kasus-kasus penyakit yang ditangani disana yang umumnya pasiennya dari daerah terdekat dari lokasi puskesmas. Pola-pola penyakit yang diderita pasien dari keseluruhan puskemas dalam regional tertentu dapat diamati untuk mengetahui pola penyakit yang berbeda yang ditangani

pada puskesmas kesalahan cara vaksinasi masyarakat (untuk anak-anak).

2.7 K-NN (*K - Nearest Neighbor*)

Seperti halnya *decision tree, K-Nearest Neighbor* sangat sering digunakan dalam klasifikasi dengan tujuan dari algoritma ini adalah untuk mengklasifikasi objek baru berdasarkan atribut dan training sample. Prinsip kerja dari K-NN (*K-Nearest Neighbor*) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan. Teknik ini termasuk dalam kelompok klasifikasi non parametrik yang dinilai lebih siap untuk dihadapkan pada berbagai kondisi data. Algoritma ini memiliki formulasi statistik paling sederhana dan paling mudah diimplementasikan. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut dan training sampel.

Algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data uji (yang klasifikasinya tidak diketahui). Jarak dan vektor yang baru ini terhadap seluruh vektor data pembelajaran dihitung, dan sejumlah K buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut. Nilai K yang terbaik untuk algoritma ini tergantung pada data secara umumnya, nilai K yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*.

2.7.1 Konsep K-NN dalam Deteksi Anomali

Tan et al (2006) mengemukakan deteksi anomali data dengan K-NN didasarkan pada adanya sejumlah data tetangga terdekat (*nearest neighbor*) dari setiap data, dari jarak K tetangga tersebut kemudian diambil jarak rata-rata dari K tetangga. Setiap data akan diberi nilai penyimpangannya atau derajat outlier. Nilai terendah derajat outlier adalah 0, dan derajat tertinggi (biasanya dinyatakan dengan jarak) bisa

mencapai tak terhingga. Definisi outlier data dalam K-NN didapatkan dari jarak pada K tetangga. Untuk data dengan derajat terendah yaitu 0 dianggap data inlier (bukan outlier), sedangkan data dengan nilai skor yang tinggi diatas T yang ditentukan, maka akan dianggap sebagai outlier.

Pemilihan jumlah tetangga yang digunakan menjadi pilihan yang sangat sensitif. Jika K terlalu kecil misal 2, maka jumlah kumpulan data yang sedikit, misal 3 (3 data tersebut terpisahkan dari kelompok yang lain) tidak akan dianggap sebagai outlier karena memiliki skor outlier yang kecil. Tetapi jika K terlalu besar maka ada kemungkinan ada sejumlah data dalam satu cluster yang jumlah datanya kurang dari K akan mempunyai skor outlier yang tinggi sehingga dianggap sebagai outlier.

Metode K-NN dalam konteks deteksi anomali menggunakan basis jarak antar data pada K tetangga terdekat, hal ini sangat sederhana untuk diimplementasikan. Tetapi pendekatan kemiripan (*proximity*) data dengan jarak mempunyai kompleksitas O(m2) kali. Hal ini akan menyebabkan biaya komputasi yang mahal untuk data set dengan ukuran besar. Pemilihan parameter jarak yang digunakan juga berpengaruh, untuk Euclidean sangat cocok untuk memberikan jarak terdekat antar data, sedangkan Manhattan biasanya lebih tangguh untuk dapat digunakan dalam mendeteksi outlier berbasis jarak. Hal tersebut disebabkan jarak Manhattan menggunakan jumlah selisih absolut antara dua buah data, sehingga mampu memberikan jarak terjauh antara dua buah data.

2.7.2 Algoritma K-NN untuk Deteksi Anomali

Berikut adalah langkah-langkah mengerjakan K-NN untuk deteksi anomali atau outlier seperti berikut :

1. Tetapkan parameter K jumlah tetangga terdekat, threshold T
2. Lakukan langkah 3 sampai 6 untuk setiap data
3. Hitung jarak ke semua data yang lain
4. Ambil K tetangga terdekat

5. Hitung rata-rata jarak dari K tetangga
6. Jika rata-rata jarak yang didapat ≥ t, tandai data tersebut sebagai outlier.

Banyak macam rumus dalam menghitung jarak, salah satunya adalah Manhattan. Manhattan ini berdasarkan pada kota Manhattan yang tersusun menjadi blok-blok. Sehingga sering juga disebut *city block distance*, juga sering disebut sebagai *ablosute value distance* atau *boxcar distance*. Sebagai contohnya, kita berjalan dari lokasi A menuju utara 3 meter, kemudian belok ke timur 4 meter. Berapakah jarak kita yang sekarang dari posisi titik A tadi? *City Block distance* adalah panjang jalan yang sudah kita tempuh dari B ke A (Prasetyo, 2012).

Penulis akan mengerjakan proposal ini dengan menggunakan hitung jarak Manhattan, karena menghitung jarak menggunakan Manhattan hasilnya akan lebih akurat dan sangat kuat jika digunakan untuk mendeteksi outlier dalam data.

$$D(x,y) = \sum_{i=1}^{N} (|x - y|) \quad \text{..... (2.1)}$$

Setelah menghitung jarak menggunakan rumus jarak Manhattan, kemudian akan dilakukan perhitungan menentukan nilai treshold terlebih dahulu untuk menentukan data yang termasuk outlier dengan rumus Standar Deviasi.

Standar deviasi adalah ukuran sebaran statistik yang paling lazim, mengukur bagaimana nilai-nilai data tersebar. Bisa juga diartikan sebagai rata-rata jarak penyimpangan titik-titik data diukur dari nilai rata-rata data tersebut. Berikut adalah rumusnya :

$$S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \quad \text{..... (2.2)}$$

Setelah mendapatkan hasil dari nilai standar deviasi, langkah selanjutnya adalah menentukan tresholdnya dengan rumus sebagai berikut :

Treshold = Mean + (3 x Standar Deviasi) ..... (2.3)

Mean diambil dari nilai rata-rata yang dihasilkan oleh rumus jarak

Manhattan, dan ditambahkan dengan hasil tiga kali standar deviasinya.

2.8 Database

Database dapat dianggap sebagai tempat sekumpulan berkas yang terkomputerisasi, pada dasarnya adalah sistem komputerisasi yang tujuan utamanya adalah melakukan pemeliharaan terhadap informasi dan membuat informasi tersebut tersedia saat dibutuhkan.

Database adalah mekanisme yang digunakan untuk menyimpan informasi atau data. Informasi adalah sesuatu yang kita gunakan sehari-hari untuk berbagai alasan. Pengguna dapat menyimpan data secara terorganisasi dengan menggunakan database. Setelah data disimpan, informasi harus mudah diambil, dan mudah ditambahkan ke dalam database, dimodifikasi, dan dihapus.

Jadi secara konsep, database atau basisdata adalah kumpulan dari banyak data yang membentuk suatu berkas yang saling berhubungan dengan tatacara yang tertentu untuk membentuk data baru atau infromasi. Database adalah kumpulan dari data yang saling berhubungan antara satu dengan yang lainnya yang diorganisasikan berdasarkan skema atau struktur tertentu. Pada komputer, database disimpan dalam perangkat hardware, dan dengan software tertentu dimanipulasi untuk kepentingan atau kegunaan tertentu. Hubungan atau relasi data biasanya ditunjukkan dengan kunci (*key*) dari tiap file yang ada.

2.9 Tinjauan Penelitian Sebelumnya

Sebagai bahan perbandingan, pada penelitian sebelumnya yang pertama dilakukan oleh Yeni Kustiyahningsih, Devie Rosa Anamisa, dan Nikmatus Syafa'ah dari Universitas Trunojoyo tentang Sistem Pendukung Keputusan Untuk Menentukan Jurusan Pada Siswa SMA Menggunakan metode K-NN dan Smart. Permasalahan yang dibahas adalah bagaimana membuat aplikasi sistem pendukung keputusan untuk membantu Guru maupun Siswa SMA dalam menentukan jurusan. Penjurusan yang ada pada siswa SMA terbagi menjadi 3 jurusan, yaitu IPA, IPS, dan BAHASA. Algoritma yang digunakan untuk menghitung nilai raport siswa digunakan metode K-NN *(K-Nearest*

Neighbor). Semua kriteria termasuk nilai raport siswa yang telah dihitung menggunakan metode K-NN akan diproses menggunakan metode pembobotan SMART. Pemilihan metode SMART dikarenakan metode ini dapat melakukan pengambilan keputusan yang multiatribut. Dalam penelitian ini ada bobot dan kriteria yang dibutuhkan untuk menentukan jurusan apa yang cocok bagi siswa tersebut, yaitu : 1) Rata-rata nilai raport semester 1 dan 2, sebanyak 12 mata pelajaran antara lain : Matematika, Fisika, Kimia, Biologi, Ekonomi, Geografi, Sejarah, Sosiologi, Bahasa Indonesia, Bahasa Inggris, Seni Budaya, dan Bahasa Jepang, 2) Hasil test psikologi, 3) Minat siswa, dan 4) Saran/anjuran orang tua. Pada penelitian ini hasil keluarannya diambil dari urutan alternatif tertinggi ke alternatif terendah. Alternatif yang dimaksud adalah jurusan IPA, IPS, atau BAHASA. Hasil akhir yang dikeluarkan oleh programnya berasal dari jumlah keseluruhan dari nilai setiap kriteria, karena dalam setiap kriteria memiliki nilai yang berbeda-beda. Hasil dari proses data mining ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan dalam menentukan jurusan yang sesuai dengan kemampuan siswa.

Penelitian yang kedua dilakukan oleh Ahmad Fajar al Kharis dari ITS, yang berjudul Deteksi Intrusi pada Jaringan Komputer Berdasarkan Analisa Payload Menggunakan Metode Outlier. Permasalahannya mendeteksi intrusi pada jaringan komputer dan mencegah serangan yang tidak diketahui dari internet berdasarkan analisa payloadnya. Penelitian ini mengenai sistem pertahanan jaringan komputer terhadap aktivitas gangguan, dengan bantuan IDS (*Intrusion Detection System*) sebuah software yang berfungsi membantu kerja admin jaringan, karena aliran data disebuah jaringan sangat banyak dan prosesnya berlangsung 24 jam. Penulis menggunakan 2 proses utama dalam penelitian ini, pertama proses picture packet pada jaringan yang berfungsi untuk mengcapture dan menuliskan data yang akan dianalisa pada proses selanjutnya, dan kedua proses analisa dari data yang sudah disimpan di file. Sebelum dianalisa program akan melakukan ekstraksi fitur karena data yang disimpan masih berbentuk binary dan tidak terbaca. Proses ekstraksi dilakukan pada paket TCP dan fokus pada bagian payloadnya, kemudian melakukan proses normalisasi terhadap fitur yang diekstrak dan menghasilkan resultan.

Resultan tersebut digunakan untuk menghitung treshold dan menentukan apakah data yang lewat termasuk serangan atau bukan. Bagian akhir program menampilkan output berupa data yang mengandung anomaly dan juga menampilkan data-data yang menyertainya seperti tanggal dan waktu kejadian alamat IP tujuan dan asal, port tujuan dan port asal, dan juga payloadnya itu sendiri. Tipe payload yang termasuk serangan bisa dicatat untuk update database serangan yang telah dikenali. Hal ini dapat membantu admin dalam mengambil keputusan selanjutnya dan juga membantu menjaga keamanan jaringan komputer. Program mampu mengenali serangan baik serangan dengan tipe lama, maupun serangan dengan tipe yang baru.