BAB II
KAJIAN PUSTAKA

2.1 *Knowledge Discovery in Database* (KDD)

Pemrosesan data secara komputerisasi telah dimplementasikan pada berbagai bidang pekerjaan seperti dunia bisnis, pendidikan maupun pemerintahan. Pemrosesan data secara komputerisasi bisa dilakukan dengan komputer dimana data disimpan dalam suatu *database*. Semakin lama data yang disimpan didalam suatu *database* akan semakin banyak dan bertumpuk sementara tidak semua data yang tersimpan digunakan secara maksimal. Hal ini menyebabkan munculnya istilah *Rich of Data but Poor of Information.*

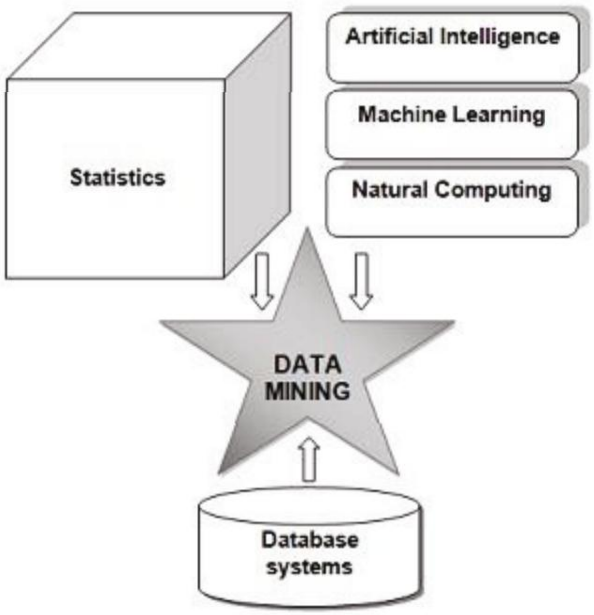
Data mining merupakan suatu solusi yang dapat menemukan pola atau pengetahuan yang bermanfaat secara otomatis dari sekumpulan data yang berjumlah banyak (Sunjana, 2010). Untuk istilah *data mining* kadang disebut juga dengan *Knowledge Discovery in Databases (KDD)*. *Knowledge Discovery in Databases* (KDD) merupakan suatu proses yang validasi, manfaat, dan pola yang bisa dipahami dalam sebuah data yang tersimpan dalam sebuah *database* yang besar (Putra, 2016).

Proses *knowledge discovery* melibatkan hasil dari proses *data mining* (proses mengekstrak kecenderungan pola suatu data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami. KDD sendiri diartikan sebagai keseluruhan proses non-trivial untuk mencari dan

mengidentifikasi pola dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti (Sari & Sindunata, 2014).

2.2 *Data Mining*

Data mining adalah suatu proses untuk menemukan hubungan, pola, dan tren baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Kamagi & Hansun, 2014). *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan (Haryati, Sudarsono, & Suryana, 2015).



Gambar 2.1 Penjelasan *Data Mining*

Dari gambar 2.1, dapat disimpulkan *data mining* merupakan suatu teknik statistik, matematik, kecerdasan buatan dan *machine learning* yang digunakan untuk menjelaskan penemuan pengetahuan didalam *database* besar.

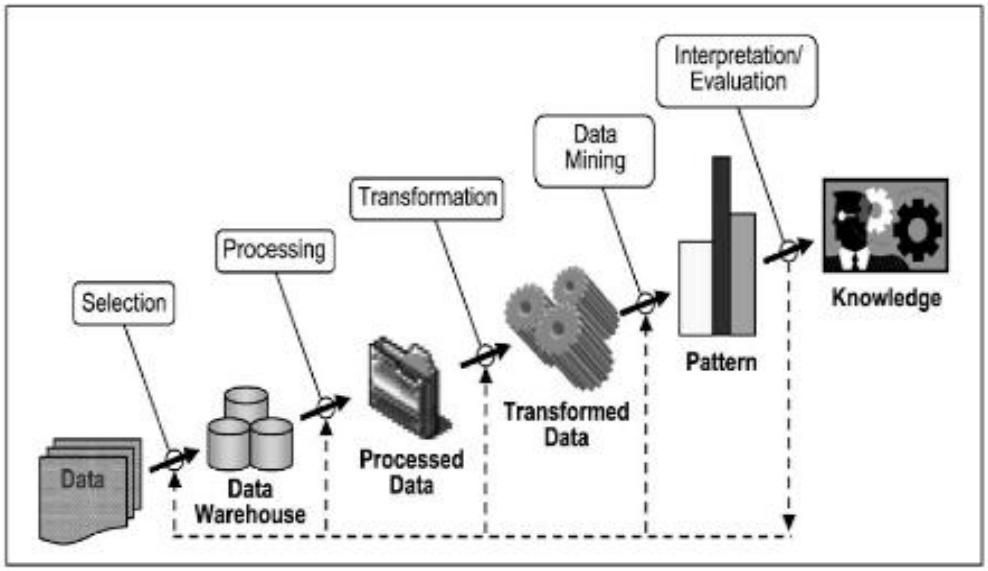
Menurut Larose (2005), terdapat beberapa faktor yang mempengaruhi kemajuan dalam bidang *data mining* antara lain (Syarif, 2015):

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam *warehouse* yang menjadikan akses ke dalam *database* yang andal.
3. Adanya peningkatan akses data melalui navigasi *web* dan intranet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Adanya ketersediaan teknologi perangkat lunak untuk *data mining*.
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Selain faktor yang mempengaruhi kemajuan *data mining*, terdapat juga kesulitan dalam mendefinisikan *data mining* karena *data mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu sehingga *data mining* ditekankan agar dapat menangani jumlah data yang sangat besar, dimensi data yang tinggi dan data yang bersifat heterogen.

2.2.1 Tahapan *Data Mining*

Tahapan *data mining* secara garis besar dapat dijelaskan sebagai berikut (Nofriansyah, 2014):



Gambar 2.2 Tahapan *Data Mining* (KDD)

Dari gambar 2.2, diketahui bahwa terdapat beberapa tahapan *data mining* yang terdiri dari:

1. Data *Selection*

Merupakan tahapan pemilihan data dari sekumpulan data di *database*. Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD (*Knowledges Discovery in Databases)* dimulai. Data yang terpilih hanyalah data yang *valid*.

2. Processing

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* atau pembersihan data yang mencakup membuang data duplikasi, memeriksa data inkonsisten, dan memperbaiki adanya tipografi pada data. Dalam

proses ini juga dilakukan proses *enrichment* yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD. Hasil dari tahap ini adalah *processed data.*

3. *Transformation*

Proses transformasi merupakan tahap pengubahan data sebelum diproses dalam *data mining*. Pengubahan data seperti proses pengolahan data dari data kuantitatif menjadi data kualitatif. Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada tujuan yang ingin dicapai. Hasil dari tahap ini adalah *transformed data* yang siap untuk dimasukkan dalam tahap *data mining.*

4. *Data mining*

Merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu seperti metode klasifikasi, asosiasi, estimasi, klustering, deskripsi dan sebagainya. Hasil dari tahap ini adalah suatu pola atau *pattern*.

5. *Interpretation/Evaluation and Presentation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini berfungsi untuk menghasilkan suatu pengetahuan baru. *Interpretation* merupakan suatu penerjemahan pola-pola yang dihasilkan dari *data mining*. Bagian dari proses ini mencakup pemeriksaan hubungan antara pola yang ditemukan dengan fakta atau hipotesa yang ada sebelumnya (Sijabat, 2015).

2.2.2 Penerapan *Data Mining*

Penerapan *data mining* dalam kehidupan sehari-hari memberikan sejumlah solusi diantaranya sebagai berikut (Santoso, 2014):

a. Mengetahui kebutuhan pasar

Data mining dapat melakukan pengelompokan akan model-model pembeli dan melakukan klasifikasi terhadap setiap pembeli sesuai dengan karakteristik seperti kesukaan dan kebiasaan belinya.

b. Melihat pola beli pemakai dari waktu ke waktu

Data mining dapat digunakan untuk melihat pola beli seseorang dari waktu ke waktu. Pola beli pada tiap orang akan berbeda dan juga untuk seseorang tertentu jika terjadi perubahan status maka pola belinya juga akan terpengaruhi. Namun besar kemungkinan pola beli pemakai sama untuk kurun waktu yang dekat.

c. *Cross-Market Analysis*

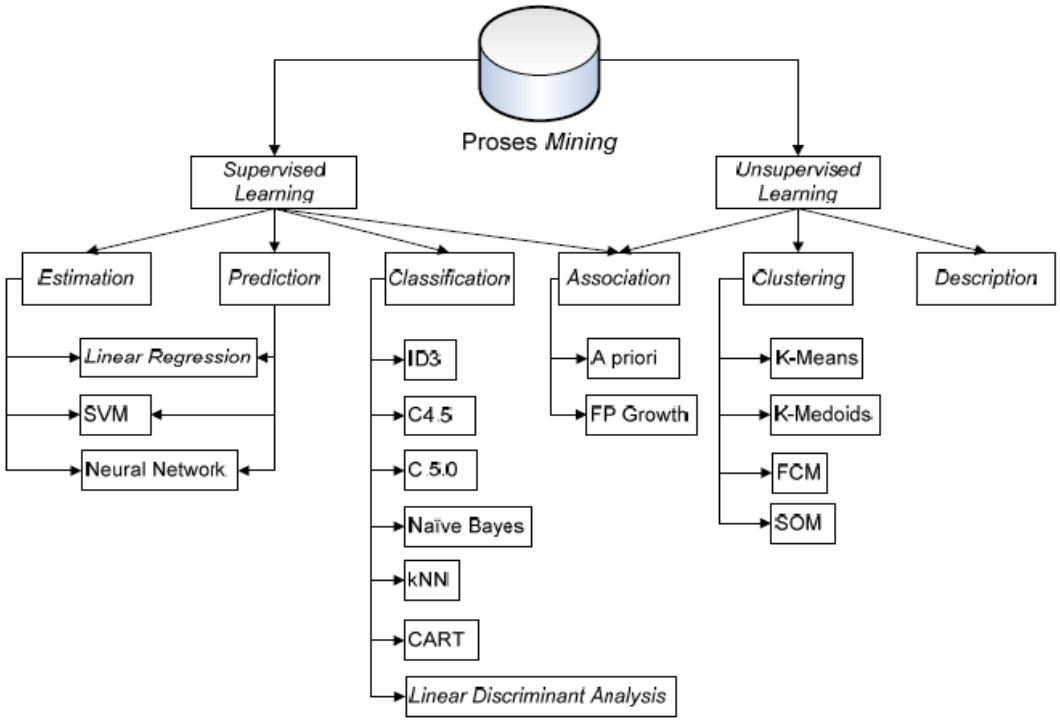
Data mining dapat digunakan untuk melihat hubungan antara penjualan satu produk dengan produk lainnya. Dengan adanya penerapan *data mining* dalam segi *cross market analysis* maka dapat dioptimasikan penjualan produk.

d. Profil *Customer*

Data mining dapat digunakan untuk memantau profil c*ustomer* sehingga dapat diketahui kesukaan atau minat pelanggan akan produk tertentu.

2.3 Metode *Data Mining*

Menurut Wu (2009), model *data mining* terdiri atas dua kelas utama sebagai berikut (Sardiarinto, 2013):



Gambar 2.3 Metode *Data Mining*

Dari gambar 2.3 tersebut, diketahui bahwa dua kelas utama tersebut adalah *supervised learning* dan *unsupervised learning* yang terdiri dari:

1. *Supervised Model*

Model *supervised* merupakan pemodelan diarahkan, bertujuan untuk memprediksi suatu kejadian atau memperkirakan nilai dari atribut numerik terus menerus.

Yang termasuk dalam model *supervised* adalah:

a. Model estimasi

Model estimasi berguna untuk memprediksi nilai dari lapangan secara kontinyu didasarkan pada nilai-nilai yang diamati dari atribut masukan. Variabel *target* model estimasi lebih kearah numerik daripada kearah kategori. Contoh algoritma yang dapat digunakan seperti algoritma *Linear Regression, Neural Network, Support Vector Machine*.

b. Model prediksi

Model prediksi berguna untuk menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang. Contoh algoritma yang dapat digunakan seperti algoritma *Linear Regression, Support Vector Machine, Neural Network*.

c. Model klasifikasi

Dalam model klasifikasi terdapat *target* variabel kategori. Model klasifikasi merupakan model kelompok atau kelas yang telah diketahui dari awal. Biasanya digunakan untuk mengelompokkan kasus ke kelompok-kelompok yang telah ditetapkan. Berbeda dengan metode *clustering* dimana variabel dependen tidak ada sedangkan pada klasifikasi, variabel dependen tersebut harus ada. Contoh algoritma yang dapat digunakan seperti algoritma ID3, C4.5, C5.0, *Naïve Bayes*, *kNN*, *CART* dan *Linear Discriminant Analysis*.

Kelas pertama pada metode *data mining* adalah model *supervised,* berikut ini adalah kelas kedua yaitu:

2. Unsupervised Model

Model *Unsupervised* dimana pengenalan pola yang terarah, tetapi tidak dipandu oleh atribut sasaran tertentu dan tidak ada bidang *output*.

a. Model asosiasi

Model asosiasi merupakan teknik *data mining* yang mempelajari hubungan antar data. Model ini tidak melibatkan prediksi langsung dari satu bidang bahkan semua bidang yang terlibat memiliki peran ganda yang bertindak sebagai masukan dan keluaran pada saat yang sama. Contoh algoritma yang dapat digunakan seperti algoritma *FP-Growth, A Priori*.

b. Model *cluster*

Model *cluster* merupakan teknik pengelompokkan data ke dalam suatu kelompok tertentu, yang kelompok tersebut tidak dikenal dari awal. Contoh algoritma yang dapat digunakan seperti algoritma *K-Means, K-Medoids, Self-Organization Map* (SOM), *Fuzzy C-Means*.

c. Model deskripsi

Model deskripsi merupakan teknik *data mining* yang digunakan untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.

2.3.1 Teknik Klasifikasi

Seiring dengan perkembangan pengetahuan *data mining* dan komponen-komponennya, *data mining* tidak lagi dimonopoli oleh bidang teknologi informasi.

Pemakainya telah semakin meluas ke bidang lain misalnya pada bidang kesehatan, pertanian, asuransi, dan lain-lain (Mardiani, 2012).

Klasifikasi data merupakan suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan objek didalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Tujuan dari klasifikasi adalah untuk menemukan model dari *training set* yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya (Ginting, Zarman, & Hamidah, 2014).

Proses klasifikasi memiliki dua tahap yaitu tahap pertama adalah proses pembelajaran dimana data pelatihan dianalisis dengan algoritma klasifikasi sehingga terbentuk aturan klasifikasi dan tahap kedua adalah proses pengujian dimana data uji digunakan untuk memperkirakan keakuratan aturan klasifikasi. Teknik klasifikasi terbagi menjadi beberapa teknik seperti *decision tree, bayesian methods, bayesian network, genetic algorithms, rough set* dan *fuzzy logic* (Huang et al., 2009).

Proses klasifikasi ditandai dengan karakteristik sebagai berikut (Gorunescu, 2011):

1. *Input* merupakan *dataset* pelatihan yang berisi objek dengan tipe atribut dan terdapat satu variabel tipe label yang berfungsi sebagai hasil.
2. *Output* merupakan sebuah model yang memberikan label khusus untuk setiap objek berdasarkan atribut lainnya.

3. *Classifier* digunakan untuk memprediksi kelas objek baru yang tidak diketahui. Sebuah *dataset* pengujian juga digunakan untuk menentukan keakuratan model.

2.3.2 Algoritma C4.5

Algoritma biasanya diartikan sebagai urutan atau langkah-langkah logis untuk menyelesaikan suatu masalah (Rifqo & Arzi, 2016). Salah satu algoritma pohon keputusan yang terkenal adalah algoritma C4.5. Algoritma C4.5 ditemukan oleh seorang peneliti di bidang mesin pembelajaran bernama *J.Ross Quinlann* yang dikembangkan dari ID3 (Evicienna & Amalia, 2013). Algoritma C4.5 memiliki prinsip yang sama dengan ID3, perbedaan utama C4.5 dari ID3 adalah:

1. C4.5 dapat menangani atribut kontinyu dan diskrit.
2. C4.5 dapat menangani *training* data dengan *missing value*.
3. Hasil pohon keputusan C4.5 akan dipangkas setelah dibentuk.
4. Pemilihan atribut yang dilakukan dengan menggunakan *gain ratio*.

Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu (Kusrini & Luthfi, 2009):

1. Menyiapkan *data training*. *Data training* yang digunakan biasanya dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atibut yang terpilih dengan cara menghitung nilai *gain* dari masing-masing atribut. Akar

pertama ditunjukkan oleh atribut dengan nilai *gain* tertinggi. Sebelum menghitung nilai *gain* dari atribut, perlu dilakukan perhitungan nilai entropi dengan rumus:

$$\text{Entropi}\ (S)\ = \sum_{i=1}^{n} -pi * log_2\, pi$$

Rumus 2.1 Perhitungan Entropi

Keterangan:

S : himpunan kasus

n : jumlah partisi S

pi : proporsi dari Si terhadap S

3. Hitung nilai *gain* dengan metode *information gain* yang berfungsi untuk mengukur efektifitas atribut dalam pengklasifikasian data, rumusnya:

$$Gain\ (\text{S}, \text{A})\ = \text{Entropi}(S) - \sum_{i=1}^{n} \frac{|\text{Si}|}{|S|} * \text{Entropi}(\text{Si})$$

Rumus 2.2 Perhitungan *Gain*

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|Si| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Perhitungan *information gain* masih memiliki sejumlah kekurangan seperti pemilihan atribut yang tidak relevan sebagai pemartisi yang terbaik pada suatu simpul. *Gain ratio* merupakan normalisasi dari *information gain* yang memperhitungkan entropi dari distribusi probabilitas subset setelah

dilakukan proses partisi (Julianto, Yunitarini, & Sophan, 2014). Secara matematis, *gain ratio* dihitung sebagai berikut:

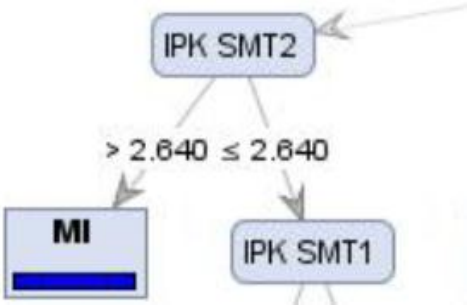
$$GainRatio(X) = \frac{Gain(X)}{SplitInfo(X)}$$

Rumus 2.3 Perhitungan *Gain Ratio*

4. Ulangi langkah ke-2 hingga semua semua tupel terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat:
   a. Semua tupel dalam *node* N mendapat kelas yang sama.
   b. Tidak ada atribut didalam tupel yang dipartisi lagi.
   c. Tidak ada tupel didalam cabang yang kosong.

2.3.3 Pohon Keputusan

Salah satu metode pengklasifikasian yang umumnya digunakan adalah pohon keputusan atau *decision tree*. *Decision tree* adalah sebuah struktur pohon, dimana setiap simpul pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan simpul daun (*leaf*) merepresentasikan kelompok kelas tertentu seperti pada gambar 2.4 (Swastina, 2013).



Gambar 2.4 Pohon Keputusan

Level simpul teratas dari sebuah *decision tree* adalah simpul akar (*root*) yang biasanya berupa atribut yang berpengaruh terbesar pada suatu kelas tertentu (Julianto et al., 2014). Adapun konsep dasar dari suatu pohon keputusan adalah mengolah data menjadi pohon keputusan, memangkas pohon keputusan dan membuat aturan keputusan.

Atribut dalam pohon keputusan digunakan untuk menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Strategi pencarian pada suatu pohon keputusan biasanya secara *top-down* sehingga dalam mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari simpul akar (*root*) sampai simpul akhir (*leaf*) dan kemudian akan diprediksi kelas yang dimiliki oleh data baru tertentu. Manfaat utama menggunakan pohon keputusan adalah kemampuannya untuk mem-*breakdown* proses pengambilan keputusan yang kompleks menjadi lebih sederhana, mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel masukan dengan sebuah variabel *target*, serta memadukan antara eksplorasi data dan pemodelan (Haryati et al., 2015).

2.3.4 *Rule Based*

The knowledge represented in decision trees can be extracted and represented in the form of classification IF-THEN rules. Pernyataan tersebut dapat diartikan bahwa pengetahuan yang diwakili dalam pohon keputusan dapat diekstraksi dan diwakili dalam bentuk aturan *IF-THEN*. Satu aturan dibuat untuk setiap jalur dari akar ke simpul daun. Bentuk *rules* atau aturan ini digunakan

untuk memudahkan manusia untuk memahaminya terutama jika pohon keputusan yang dihasilkan sangat besar (Kargupta, Han, Yu, Motwani, & Kumar, 2008).

Han & Kamber (2006) menjelaskan bahwa *rule based* atau algoritma berbasis aturan merupakan cara terbaik untuk merepresentasikan sejumlah bit data atau pengetahuan yang biasa dituliskan dalam bentuk logika *if-then* (Haryati et al., 2015). Dalam logika *if-then* tersebut terdapat pernyataan *if* yang dikenal sebagai rule *antecedent* atau *precondition* sedangkan pernyataan *then* yang disebut sebagai *rule consequent*. Adapun pada tiap pernyataan terdapat satu atau lebih atribut dengan penghubung yaitu logika *AND*. Pada *rule based* tidak digunakan logika *OR*, hal ini dikarenakan aturan-aturan diekstraksi langsung dari pohon keputusan yang disebut *mutually exclusive* (tidak adanya benturan) dan *exhaustive* (kombinasi nilai yang mungkin).

2.4 *Software* Pendukung

Data mining dapat diolah dengan menggunakan perangkat lunak aplikasi seperti WEKA, *Orange, Microsoft Analysis Services, Oracle Data Mining* dan *RapidMiner*. Pada penelitian ini, penulis akan menggunakan *software RapidMiner* untuk pengujian terhadap pohon keputusan. *RapidMiner* adalah suatu perangkat lunak aplikasi yang diciptakan oleh *Dr. Markus Hofmann* dari *Institute of Technology Blanchardstown* dan *Ralf Klinkenberg* dari *rapid-i.com* (Haryati et al., 2015). *RapidMiner* dikembangkan dari bahasa pemrograman *java* di bawah lisensi *GNU Public Licences*. *RapidMiner* memiliki sistem yang komprehensif untuk

analisa data serta banyak digunakan karena kemampuan, fleksibilitas dan kemudahan dalam penggunaannya (Purwaningsih, 2016). Kemampuan, fleksibilitas dan kemudahan *RapidMiner* dikarenakan adanya tampilan GUI (*Graphical User Interface*), bersifat *open-source* dan multi-platform.

RapidMiner is an open source data mining tool that provides data mining and machine learning procedures including data loading and transformation, data preprocessing and visualization, modelling, evaluation, and deployment (Adhatrao, Gaykar, Dhawan, Jha, & Honrao, 2013). Pernyataan diatas diterjemahkan sebagai *RapidMiner* merupakan suatu program *data mining* yang bersifat *open source* yang menyediakan perhitungan *data mining* dan prosedur *machine learning* meliputi *data loading* dan transformasi, *data preprocessing and visualization*, pemodelan, evaluasi dan pengembangan. *RapidMiner* memberikan fasilitas yang lengkap kepada pengguna *data mining*. Model yang disediakan meliputi model *Bayesian, Modelling, Tree Induction, Neural Network* dan masih banyak lagi. Sementara untuk metode yang disediakan oleh *RapidMiner* seperti klasifikasi, *clustering*, asosiasi dan sebagainya. Adapun keunggulan lain dari *RapidMiner* dimana program bersifat *open-source* sehingga jika terdapat model yang tidak tersedia maka dapat ditambahkan oleh pengguna.

2.5 *Security*

Security atau satpam adalah satuan kelompok petugas yang dibentuk oleh instansi, proyek, badan usaha untuk menyelenggarakan keamanan di lingkungan

kerjanya. Pengertian "satuan kelompok petugas" adalah satpam yang bertugas menempati pos penjagaan seorang diri maupun berkelompok. Para personil *security* bertugas melindungi harta benda melalui keberadaan dirinya dengan tingkat visibilitas yang tinggi untuk mencegah tindakan-tindakan kriminal seperti pencurian, kebakaran, maupun gangguan keamanan lainnya di lingkungan kerjanya, baik dengan pengamatan secara langsung dengan cara berpatroli, mengecek sistem alarm dan CCTV untuk kemudian mengambil tindakan dan melaporkan setiap kejadian kepada pimpinan. Untuk dapat menjalankan tugasnya dengan baik, anggota *security* yang profesional wajib memahami tugas pokok, fungsi dan peranan satuan pengamanan (TUPOKSIRAN-SATPAM).

Sesuai dengan Peraturan Kepala Kepolisian Negara Republik Indonesia nomor 24 tahun 2007, bahwa *security* sebagai bentuk pengamanan yang bertugas membantu polri di bidang penyelenggaraan keamanan dan ketertiban masyarakat, terbatas pada lingkungan kerjanya (Republik Indonesia, 2007). Tugas pokok *security* seperti menyelenggarakan keamanan dan ketertiban di lingkungan atau tempat kerjanya yang meliputi aspek pengamanan fisik, personil, informasi dan pengamanan teknis lainnya. Sementara itu, fungsi anggota *security* adalah melindungi dan mengayomi lingkungan atau tempat kerjanya dari setiap gangguan keamanan, serta menegakkan peraturan dan tata tertib yang berlaku di lingkungan kerjanya.

Faktor dasar pembentukan anggota *security* yang profesional meliputi perilaku (*attitude*), keterampilan (*knowledge*) dan pengetahuan (*skill*). Ketiga faktor tersebut harus dimiliki setiap anggota *security* mulai dari tahap perekrutan,

pelatihan hingga penempatan tugas. Adapun beberapa-beberapa syarat umum untuk menjadi anggota *security* sebagai berikut:

1. WNI

Untuk menjadi anggota *security* harus merupakan Warga Negara Indonesia (WNI) dan ini berlaku untuk semua instansi pemerintah dan perorangan sesuai ketentuan.

2. Tes Sehat

Tes Sehat menjadi suatu syarat yang penting karena personil *security* harus memiliki fisik yang kuat dan sehat untuk dapat menjamin keamanan lingkungan kerjanya.

3. Tingkat pendidikan

Pendidikan formal terakhir seorang kandidat *security* adalah minimal SMU atau sederajat, dengan postur tinggi badan minimal 165 cm untuk pria dan minimal 160 cm untuk wanita dengan berat badan yang proporsional. Hal ini sesuai ketentuan yang berlaku untuk dapat menunjang kerja seorang *security* dan menunjukan kewibawaan sebagai seorang personil *security*.

4. Usia

Seorang calon personil *security* minimal berusia 20 tahun dan maksimal 30 tahun, walaupun persyaratan tersebut terkadang tidak diterapkan oleh beberapa perusahaan, karena ada juga pensiunan TNI dan polisi yang menjadi satpam karena menyesuaikan kebutuhan perusahaan dalam hal kondisi tertentu di lapangan.

5. Psikotes

Selain fisik yang kuat dan sehat, seorang satpam juga diharapkan memiliki karakter, mental serta moral yang baik, karena berhubungan dengan tugas menegakkan peraturan yang berlaku.

6. Bebas narkoba

Seorang personil *security* harus terbebas dari narkoba dan untuk memperoleh surat keterangan bebas narkoba, perusahaan atau instansi para pelamar calon *security* juga menunjuk rumah sakit setempat untuk melakukan tes bebas narkoba.

Setiap personil *security* wajib memiliki kompetensi yang baik dalam pelaksanaan tugasnya. Tempat pendidikan dan pelatihan *security* adalah melalui Lembaga Pendidikan Kepolisian Negara ataupun melalui Badan Usaha Jasa Pengamanan (BUJP) yang sudah memiliki izin operasional dari kepolisian. Setiap peserta yang telah lulus pendidikan *security* akan dibekali dengan KTA (Kartu Tanda Anggota) maupun sertifikat dari Polda setempat.

Sebagaimana diatur oleh Peraturan Kapolri Nomor 18 Tahun 2006 tentang Pelatihan dan Kurikulum Satuan Pengamanan, jenjang pelatihan satpam terdiri dari 3 tingkatan yaitu (Republik Indonesia, 2006):

1. Dasar (Gada Pratama)

Pelatihan ini merupakan dasar dan wajib bagi calon anggota *security*. Lama pelatihan empat minggu dengan pola 232 jam pelajaran. Materi pelatihan antara lain: Kemampuan Interpersonal, Etika Profesi, Tugas Pokok, Fungsi dan Peranan Satpam, Wewenang Kepolisian Terbatas, Bela Diri, Pengenalan Bahan Peledak,

Pengetahuan tentang Narkotika, Psikotropika dan Zat Adiktif Lainnya, Penggunaan Tongkat Polri dan Borgol, Pengetahuan Baris Berbaris dan Penghormatan.

2. Penyelia (Gada Madya)

Merupakan pelatihan lanjutan bagi anggota *security* yang telah memiliki kualifikasi Gada Pratama. Lama pelatihan dua minggu dengan pola 160 jam pelajaran.

3. Manajer Keamanan (Gada Utama)

Pelatihan ini merupakan pelatihan yang boleh diikuti oleh siapa saja dalam level setingkat manajer yaitu kepala keamanan atau manajer keamanan dengan pola pelatihan adalah 100 jam pelajaran.

2.6 Penelitian Terdahulu

Ada beberapa penelitian-penelitian terkait yang dibuat oleh beberapa peneliti sebelumnya yaitu:

1. Penelitian yang dilakukan oleh (Rajesh & Anand, 2012) yang berjudul ***ANALYSIS OF SEER DATASET FOR BREAST CANCER DIAGNOSIS USING C4.5 CLASSIFICATION ALGORITHM***, membahas tentang bagaimana memprediksi ada tidaknya kanker dan membedakan antara kasus ganas dan jinak. Pengelompokkan data kanker payudara dilakukan dengan menggunakan algoritma C4.5. Variabel masukan adalah *age at diagnosis, regional nodes positive, sequence number, cs tumor size, cs extension* dan

variabel keluaran berupa *behavior* dengan kategori *malignant potential* atau *carcinoma in situ.* Penelitian mengambil 500 data sampel secara acak dan menghasilkan tingkat akurasi sebesar 94% pada tahap pelatihan dan tingkat akurasi sebesar 93% pada tahap pengujian.

2. Penelitian yang dilakukan oleh (Florence.T & R, 2013) berjudul ***TALENT KNOWLEDGE ACQUISITION USING C4.5 CLASSIFICATION ALGORITHM***, membahas tentang pembangunan aturan klasifikasi untuk memprediksi potensi bakat yang membantu dalam menentukan apakah individu merupakan orang yang berbakat atau tidak. Variabel-variabel yang digunakan untuk penelitian seperti variabel ID, kategori, kualifikasi, efisiensi, pengalaman, kualifikasi teknik, nilai evaluasi, keahlian bahasa pemrograman C, keahlian bahasa pemrograman *Java*, keahlian bahasa pemrograman .*Net*, keahlian ITES sementara variabel keluaran berupa rekomendasi anggota berbakat. Hasil penelitian dengan pengambilan 200 sampel data secara acak mencapai tingkat akurasi sebesar 98% yang mana hanya terdapat 4 data yang hasilnya ambigu.
3. Penelitian yang dilakukan oleh (Santoso, 2014) berjudul **ANALISA DAN PENERAPAN METODE C4.5 UNTUK PREDIKSI LOYALITAS PELANGGAN**, masalah penelitiannya adalah bagaimana potensi algoritma C4.5 untuk memprediksi loyalitas pelanggan. Variabel yang digunakan adalah usia, pelayanan, promosi, harga, citra perusahaan dan kepercayaan. Hasil penelitiannya membuktikan bahwa analisa penggunaan *data mining* dengan algorima C4.5 dapat digunakan pada *dataset* pelanggan ke dalam

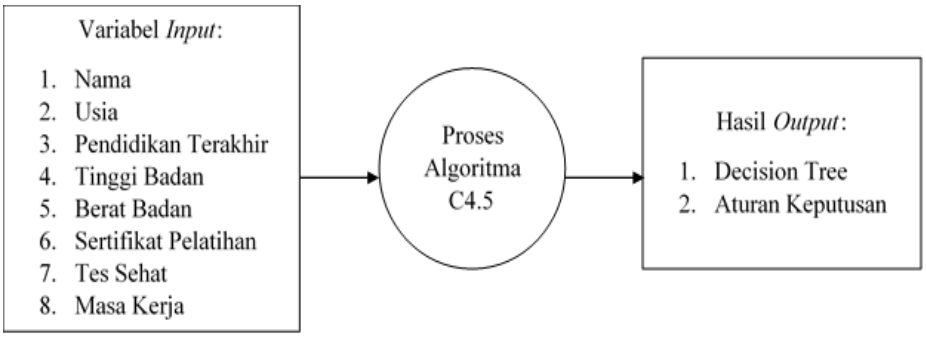
kegiatan manajemen strategi sehingga dapat mempertahankan loyalitas pelanggannya.

4. Penelitian yang dilakukan oleh (Haryati et al., 2015) yang berjudul **IMPLEMENTASI *DATA MINING* UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS: UNIVERSITAS DEHASEN BENGKULU)**, dengan permasalahan bagaimana implementasi algoritma C4.5 dalam membentuk suatu perancangan sistem baru untuk menggantikan sistem lama yang berfungsi untuk menganalisa prediksi kelulusan mahasiswa menggunakan *software RapidMiner*. Hasil yang diperoleh adalah suatu sistem baru yang lebih mudah, efektif dan efisien.
5. Penelitian yang dilakukan oleh (Setiadi, 2015) yang berjudul **PENERAPAN ALGORITMA *DECISION TREE* C4.5 UNTUK PENILAIAN RUMAH TINGGAL**, membahas tentang penggunaan *data mining* metode klasifikasi untuk penilaian rumah tinggal pengajuan kredit khususnya rumah tinggal berupa tanah dan bangunan. Pengolahan data dengan menggunakan *software RapidMiner* pada objek penelitian yang diambil dari satu set model simulasi berdasarkan data nyata yang didapatkan dari lembaga penilai kredit. Hasil pengolahan menunjukkan bahwa algoritma C4.5 dapat diterapkan dalam penilaian agunan kredit dengan akurasi 72,73% yang lebih baik dibandingkan dengan algoritma lain seperti k-NN dengan akurasi 58.82% dan Naïve Bayes dengan akurasi 54,55%.

6. Penelitian yang dilakukan oleh (Putra, 2016) yang berjudul **ALGORITMA C4.5 UNTUK MENENTUKAN TINGKAT KELAYAKAN MOTOR BEKAS YANG AKAN DIJUAL**, masalah penelitiannya adalah bagaimana *data mining* algoritma C4.5 dapat diterapkan untuk menentukan kelayakan motor bekas yang akan dijual dan bagaimana *data mining* membantu Sun Motor dalam pengambilan keputusan dalam pengadaan motor-motor yang akan dijual. Variabel penelitian meliputi mesin, rangka, *body*, cat, aki dan aksesoris. Hasil penelitian disimpulkan bahwa pemilihan variabel mempengaruhi *rule* atau *knowledge* yang dihasilkan dan algoritma C4.5 cocok diterapkan untuk mengambil keputusan terbaik dalam menentukan tingkat kelayakan motor bekas yang akan dijual.
7. Penelitian yang dilakukan oleh (Purwaningsih, 2016) yang berjudul **SELEKSI MOBIL BERDASARKAN FITUR DENGAN KOMPARASI METODE KLASIFIKASI *NEURAL NETWORK, SUPPORT VECTOR MACHINE*, DAN ALGORITMA C4.5**, membahas tentang klasifikasi seleksi mobil merk Daihatsu di PT. Tunas Mobilindo Perkasa dengan menggunakan *RapidMiner* sebagai *software* untuk menganalisa perbandingan antar metode yang ada yakni metode *Neural Network, Support Vector Machine* dan algoritma C4.5. Hasil yang diperoleh adalah model algoritma C4.5 memiliki nilai akurasi paling tinggi yaitu 82,96% dibandingkan model *Neural Network* sebesar 82,11% dan *Support Vector Machine* sebesar 76,20%.

2.7 Kerangka Pemikiran

Dalam penelitian ini, penulis akan menggunakan skema atau *flowchart* berikut:



Gambar 2.5 Kerangka Pemikiran

Penjelasan dari gambar 2.5, bahwa awal proses penelitian adalah data-data dalam gudang data diproses melalui tahapan *data mining* seperti *data selection, data cleaning*, transformasi data, *data mining, interpretation* untuk membentuk data penelitian dalam menentukan kelayakan anggota *security.* Dengan adanya penyeleksian terbentuk suatu data penelitian dari gudang data yang ada. Variabel *input* yang diambil dari data penelitian terdiri dari variabel nama, usia, pendidikan terakhir, tinggi badan, berat badan, sertifikat pelatihan, tes sehat dan masa kerja. Selanjutnya data penelitian diproses dengan menggunakan algoritma C4.5, dan hasil dari proses tersebut adalah sebuah pohon keputusan dan aturan keputusan untuk menentukan kelayakan anggota *security.*