

**PENERAPAN *DATA MINING* UNTUK MENENTUKAN
KELAYAKAN ANGGOTA *SECURITY*
DENGAN ALGORITMA C4.5**

SKRIPSI



**Oleh:
Selvina
140210060**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2018**

**PENERAPAN *DATA MINING* UNTUK MENENTUKAN
KELAYAKAN ANGGOTA *SECURITY*
DENGAN ALGORITMA C4.5**

SKRIPSI

**Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**



**Oleh:
Selvina
140210060**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2018**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Selvina

NPM : 140210060

Fakultas : Teknik dan Komputer

Program Studi : Teknik Informatika

Menyatakan bahwa Skripsi yang saya buat dengan judul:

“PENERAPAN *DATA MINING* UNTUK MENENTUKAN KELAYAKAN ANGGOTA *SECURITY* DENGAN ALGORITMA C4.5”

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip di dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah skripsi ini digugurkan dan gelar saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 7 Februari 2018



Selvina

140210060

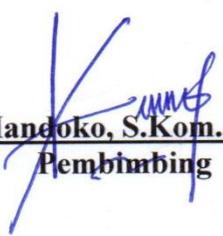
**PENERAPAN *DATA MINING* UNTUK MENENTUKAN
KELAYAKAN ANGGOTA *SECURITY*
DENGAN ALGORITMA C4.5**

**Oleh:
Selvina
140210060**

**SKRIPSI
Untuk memenuhi salah satu syarat
guna memperoleh gelar Sarjana**

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera di bawah ini**

Batam, 7 Februari 2018


**Koko Handoko, S.Kom., M.Kom.
Pembimbing**

ABSTRAK

Pentingnya jasa *security* untuk menunjang keamanan lingkungan di Indonesia masih kurang disadari oleh masyarakat. Hal tersebut dapat dilihat dari banyaknya anggota *security* yang tidak memenuhi kriteria atau syarat wajib seorang *security*. Masih banyak perekrutan anggota *security* yang dilakukan secara sembarang bahkan ada yang menggunakan unsur kekeluargaan. Penelitian ini bertujuan menentukan kelayakan anggota *security* dengan penerapan *data mining*. *Data mining* merupakan suatu proses untuk menemukan informasi-informasi yang bermanfaat dari sekumpulan *database* yang berukuran besar. Salah satu teknik yang ada pada *data mining* adalah klasifikasi. Dengan menerapkan teknik klasifikasi diharapkan dapat menghasilkan suatu pohon keputusan yang berfungsi untuk membantu direktur dan manajer dalam menentukan kelayakan anggota *security*. Adapun dalam membentuk pohon keputusan ini memerlukan suatu metode algoritma berupa algoritma C4.5. Pemodelan algoritma C4.5 dilakukan dengan perhitungan nilai entropi, *gain*, *split info* dan *gain ratio* dari data penelitian. Data penelitian bersumber dari PT Avava Duta Indonesia, data yang diambil meliputi data nama, usia, pendidikan terakhir, berat badan, tinggi badan, tes sehat, sertifikat pelatihan, dan masa kerja. Hasil penelitian adalah sebuah pohon keputusan dengan variabel-variabel yang mempengaruhi penentuan kelayakan dengan hasil akhir layak atau tidak layaknya anggota *security*. Pohon keputusan yang dihasilkan dievaluasi dengan *software RapidMiner* dan dibuktikan benar karena hasilnya sama. Berdasarkan hasil penelitian maka disimpulkan bahwa penerapan *data mining* cocok digunakan untuk menentukan kelayakan anggota *security*. Pengujian terhadap hasil penelitian juga dilakukan dengan menggunakan metode *split validation* yang menghasilkan tingkat akurasi 90,91%, *cross k-fold validation* 10 dan 3 dengan tingkat akurasi 76,32% dan 86,34%, serta *cross leave one out validation* dengan tingkat akurasi 84,21%.

Kata kunci: *Data Mining, Klasifikasi, Algoritma C4.5, Pohon Keputusan.*

ABSTRACT

The importance of security services to support environmental security in Indonesia is still poorly realized by the public. This can be seen from the number of security members who do not have the criteria or mandatory requirements of a security guard. There are still many recruitment of security members who do arbitrarily and some even use the elements of kinship. This research aims to determine the eligibility of security members with the application of data mining. Data mining is a process for finding useful information from a large set of databases. One of the techniques that exist in data mining is classification. By applying classification techniques, the research is expected to generate a decision tree that can serve to assist the director and manager in determining the eligibility of security members. As for forming a decision tree requires an algorithmic method of C4.5. Modeling by using C4.5 algorithm is done by calculation of entropy, gain, split information and gain ratio from research data. The research data is taken from PT Avava Duta Indonesia which includes data of name, age, last education, weight, height, healthy test, training certificate, and years of service. The result of research is a decision tree with the variables that influence in the determination of eligibility with the final result whether the particular member is eligible or not. The resulting decision tree is evaluated with RapidMiner and is proved true because the result is the same. Based on the results of the research, it's concluded that the application of data mining is suitable to determine the eligibility of security members. Tests on the results of research are also conducted by using split validation method with 90.91% accuracy, cross k-fold validation 10 and 3 with 76.32% and 86.34% accuracy, and cross leave one out validation with 84.21% accuracy.

Keywords: *Data Mining, Classification, C4.5 Algorithm, Decision Tree.*

KATA PENGANTAR

Syukur terhadap Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati.

Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam.
2. Dekan Fakultas Teknik dan Komputer Universitas Putera Batam.
3. Ketua Program Studi Teknik Informatika Universitas Putera Batam.
4. Bapak Koko Handoko, S.Kom., M.Kom. selaku pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
5. Dosen dan Staff Universitas Putera Batam.
6. Bapak Ahmad Arifin selaku Direktur PT Avava Duta Indonesia yang telah memberikan izin pengumpulan data penelitian.
7. Kedua orang tua dan anggota keluarga yang selalu mendoakan dan memberi semangat.
8. Teman-teman perkuliahan yang telah membantu dan memberikan informasi yang berguna.

Semoga Tuhan Yang Maha Esa membalas kebaikan dan selalu mencurahkan hidayat serta taufikNya, Amin.

Batam, Februari 2018

Penulis

DAFTAR ISI

Halaman

HALAMAN SAMPUL DEPAN	
HALAMAN JUDUL	ii
SURAT PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	3
1.3 Pembatasan Masalah	4
1.4 Perumusan Masalah	4
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
1.6.1 Manfaat Teoritis.....	5
1.6.2 Manfaat Praktis	6
BAB II KAJIAN PUSTAKA	
2.1 <i>Knowledge Discovery in Database (KDD)</i>	7
2.2 <i>Data Mining</i>	8
2.2.1 Tahapan <i>Data Mining</i>	10
2.2.2 Penerapan <i>Data Mining</i>	12
2.3 Metode <i>Data Mining</i>	13
2.3.1 Teknik Klasifikasi.....	15
2.3.2 Algoritma C4.5	17
2.3.3 Pohon Keputusan	19
2.3.4 <i>Rule Based</i>	20
2.4 <i>Software</i> Pendukung	21
2.5 <i>Security</i>	22
2.6 Penelitian Terdahulu	26
2.7 Kerangka Pemikiran.....	30
BAB III METODE PENELITIAN	
3.1 Desain Penelitian	31
3.2 Teknik Pengumpulan Data.....	34
3.3 Operasional Variabel	34

3.4	Metode Analisis dan Rancangan Sistem.....	37
3.5	Lokasi dan Jadwal Penelitian.....	38
3.5.1	Lokasi Penelitian.....	38
3.5.2	Jadwal Penelitian	39

BAB IV HASIL DAN PEMBAHASAN

4.1	Analisa Data.....	40
4.1.1	Seleksi Data	40
4.1.2	<i>Cleaning</i> Data	41
4.1.3	Transformasi Data.....	42
4.1.4	Perhitungan Algoritma C4.5	45
4.2	Hasil Pengujian	64
4.2.1	Implementasi <i>RapidMiner</i>	64
4.2.2	Pengujian dengan <i>Split Validation</i>	70
4.2.3	Pengujian dengan <i>Cross Validation (K-Fold Validation)</i>	71
4.2.4	Pengujian dengan <i>Cross Validation (Leave One Out)</i>	72

BAB V SIMPULAN DAN SARAN

5.1	Simpulan	74
5.2	Saran	75

DAFTAR PUSTAKA	76
-----------------------------	-----------

DAFTAR RIWAYAT HIDUP

SURAT KETERANGAN PENELITIAN

LAMPIRAN

DAFTAR TABEL

	Halaman
Tabel 3.1 Data Karyawan PT Avava Duta Indonesia	36
Tabel 3.2 Jadwal Penelitian.....	39
Tabel 4.1 Perhitungan Kategori Sampel.....	41
Tabel 4.2 Klasifikasi Atribut Usia.....	43
Tabel 4.3 Klasifikasi Tinggi Badan.....	43
Tabel 4.4 Klasifikasi Berat Badan.....	44
Tabel 4.5 Klasifikasi Tes Kesehatan	44
Tabel 4.6 Klasifikasi Pendidikan Terakhir.....	44
Tabel 4.7 Klasifikasi Sertifikat Pelatihan.....	45
Tabel 4.8 Klasifikasi Masa Kerja	45
Tabel 4.9 Sampel Penelitian.....	46
Tabel 4.10 Hasil Perhitungan <i>Node</i> 1.....	55
Tabel 4.11 Hasil Perhitungan <i>Node</i> 2.....	57
Tabel 4.12 Hasil Perhitungan <i>Node</i> 3.....	58
Tabel 4.13 Hasil Perhitungan <i>Node</i> 4.....	60
Tabel 4.14 Hasil Perhitungan <i>Node</i> 5.....	61
Tabel 4.15 Hasil Perhitungan <i>Node</i> 6.....	63

DAFTAR GAMBAR

	Halaman
Gambar 2.1	Penjelasan <i>Data Mining</i> 8
Gambar 2.2	Tahapan <i>Data Mining</i> (KDD) 10
Gambar 2.3	Metode <i>Data Mining</i> 13
Gambar 2.4	Pohon Keputusan 19
Gambar 2.5	Kerangka Pemikiran 30
Gambar 3.1	Desain Penelitian 31
Gambar 3.2	Rancangan Sistem 37
Gambar 4.1	Pohon Keputusan <i>Node</i> Akar 56
Gambar 4.2	Pohon Keputusan <i>Node</i> 2 58
Gambar 4.3	Pohon Keputusan <i>Node</i> 3 59
Gambar 4.4	Pohon Keputusan <i>Node</i> 4 61
Gambar 4.5	Pohon Keputusan <i>Node</i> 5 62
Gambar 4.6	Pohon Keputusan <i>Node</i> 6 63
Gambar 4.7	Gambaran Hubungan Proses 65
Gambar 4.8	Paramater <i>Read Excel</i> 65
Gambar 4.9	<i>Import Data ke RapidMiner</i> 66
Gambar 4.10	Pemilihan Variabel 66
Gambar 4.11	Bagian Dalam Proses <i>Validation</i> 67
Gambar 4.12	Pohon Keputusan <i>RapidMiner</i> 67
Gambar 4.13	Hasil Pengujian <i>Split Validation</i> 70
Gambar 4.14	Hasil Pengujian <i>K-Fold Validation 10</i> 71
Gambar 4.15	Hasil Pengujian <i>K-Fold Validation 3</i> 72
Gambar 4.16	Hasil Pengujian <i>Leave One Out Validation</i> 72

DAFTAR RUMUS

	Halaman
Rumus 2.1 Perhitungan Entropi	18
Rumus 2.2 Perhitungan <i>Gain</i>	18
Rumus 2.3 Perhitungan <i>Gain Ratio</i>	19
Rumus 4.1 Perhitungan Sampel	40
Rumus 4.2 Perhitungan Sampel <i>Stratified Random</i>	41
Rumus 4.3 Perhitungan Nilai Interval.....	42
Rumus 4.4 Perhitungan <i>Body Mass Index</i>	43
Rumus 4.5 Perhitungan Akurasi	70

DAFTAR LAMPIRAN

- Lampiran 1. Data-data Penelitian
- Lampiran 2. Hasil Transformasi Data
- Lampiran 3. *Dataset* dalam *RapidMiner*
- Lampiran 4. Perhitungan *Node 2*
- Lampiran 5. Perhitungan *Node 3*
- Lampiran 6. Perhitungan *Node 4*
- Lampiran 7. Perhitungan *Node 5*
- Lampiran 8. Perhitungan *Node 6*

BAB I PENDAHULUAN

1.1 Latar Belakang

Pada zaman sekarang ini, tindakan kriminal semakin bertambah. Dilihat dari data jumlah tindak pidana menurut kepolisian daerah yang diambil dari data Badan Pusat Statistik ditunjukkan bahwa pada tahun 2015, angka tindak pidana di Kepri mencapai 4.892 (Sub Direktorat Statistik Politik dan Keamanan, 2016). Ancaman dan risiko tersebut sangatlah beragam baik itu kasus kecil yang hanya merugikan pribadi tertentu hingga kasus besar yang merugikan perusahaan. Tindakan kriminal pada perusahaan harus diantisipasi sehingga tidak membawakan pengaruh buruk terhadap karyawan, fasilitas dan aset perusahaan. Oleh karena itu, penggunaan jasa *security* sebagai suatu kebutuhan bagi perusahaan untuk menjamin keamanan dan ketertiban.

Dunia *security* satpam di Indonesia memiliki masalah yang sangat kompleks, terkait sumber daya manusia, manajemen dan kompleksitas teknologi keamanan. Hingga saat ini, anggota *security* masih belum dikelola dengan baik dan maksimal meskipun mereka memainkan peranan yang penting dalam hal perlindungan dan pengamanan. Masyarakat umum masih menganggap remeh terhadap profesi tersebut. Hal ini terjadi akibat kurangnya kesadaran masyarakat tentang keamanan. Setiap *security* harus memiliki kompetensi dan dalam pelaksanaan tugasnya,

seorang *security* juga harus senantiasa mematuhi kode etik satpam dan prinsip penuntun satpam (Republik Indonesia, 2006).

Pemahaman manusia mengenai keamanan tidak seluruhnya salah dikarenakan dalam keseharian, masih banyak ditemukan anggota *security* yang kurang berkompeten baik itu yang bertugas di mal, sekolah maupun pabrik. Hal tersebut dikarenakan adanya perekrutan anggota *security* yang kurang memperhatikan aspek wajib atau standar pada seorang *security*. Perekrutan dilakukan secara sembarang bahkan ada yang menggunakan unsur kekeluargaan.

Anggota *security* yang tidak memiliki kecakapan yang tinggi tentunya juga membawakan dampak terhadap reputasi instansi atau perusahaan penyedia jasa *security*. Untuk menghindari hal tersebut, perlu adanya pertimbangan oleh lembaga penyedia jasa *security* dalam hal perekrutan dan penentuan anggota *security* yang layak. Adapun tingkat kelayakan anggota *security* dapat dinilai melalui fisik personil, kinerja personil, prestasi anggota, penanganan kejadian dan kelengkapan peralatan kerja.

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam *database* (Pradana, Ubaidillah, & Alfiyah, 2014). Teknologi *data mining* mulai muncul karena akumulasi data yang besar dan pesat pertumbuhannya sehingga menimbulkan istilah *rich of data but poor information*. Penilaian terhadap kelayakan anggota *security* dapat dilakukan dengan menerapkan *data mining*. Adapun *data mining* memiliki beberapa metode, metode yang akan diterapkan adalah metode klasifikasi menggunakan algoritma C4.5. Metode klasifikasi merupakan model yang kelompok atau kelas diketahui dari

awal dan digolongkan ke kelompok-kelompok yang telah ditetapkan (Sardiarinto, 2013). Algoritma C4.5 berfungsi untuk mengklasifikasikan data dalam bentuk tingkatan yang berbeda mulai dari akar ke daun secara hirarki. Proses tersebut akan membentuk pohon keputusan. Dari pohon keputusan dapat dihasilkan *rules* atau aturan-aturan yang memudahkan untuk pemahaman terhadap pohon yang berskala besar.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah diatas, dapat diidentifikasi permasalahan-permasalahan sebagai berikut:

1. Banyaknya tindakan kriminal sehingga perusahaan klien perlu menggunakan jasa *security* dan jasa *security* yang disediakan harus dapat bersaing dan memenuhi tuntutan dari perusahaan klien.
2. Perekrutan anggota *security* yang dilakukan seadanya bahkan dengan unsur kekeluargaan tanpa memperhatikan standar yang ada sehingga ada anggota *security* masih dinilai tidak layak.
3. Anggota *security* yang tidak memiliki kecakapan yang tinggi membawakan dampak terhadap reputasi instansi atau perusahaan penyedia jasa *security* yang menyebabkan kurangnya kepercayaan perusahaan klien.

1.3 Pembatasan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang telah diuraikan diatas, maka perlu adanya pembatasan masalah agar pembahasan ini tetap terfokus dan sesuai dengan tujuan yang akan dicapai, pembatasan masalah penelitian sebagai berikut:

1. Data penelitian yang digunakan bersumber dari PT Avava Duta Indonesia sebagai salah satu perusahaan penyedia jasa *security* di Batam.
2. Kriteria yang digunakan sebagai variabel dalam menentukan kelayakan anggota *security* adalah nama, usia, pendidikan terakhir, tinggi badan, berat badan, sertifikat pelatihan, tes sehat dan masa kerja.
3. Metode yang digunakan adalah metode klasifikasi dengan menggunakan algoritma C4.5 untuk membuat pohon keputusan.
4. Aplikasi yang digunakan untuk menguji hasil penelitian adalah *RapidMiner*.
5. Hasil dari penelitian hanya menganalisis layak atau tidak layaknya anggota *security* tanpa membangun *software* aplikasi.

1.4 Perumusan Masalah

Berdasarkan masalah diatas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan *data mining* algoritma C4.5 untuk menentukan kelayakan anggota *security*?

2. Bagaimana analisis tingkat akurasi untuk hasil perhitungan dengan menggunakan *software RapidMiner*?

1.5 Tujuan Penelitian

Sesuai dengan perumusan masalah diatas, maka tujuan penelitian ini adalah sebagai berikut:

1. Untuk mendeskripsikan penerapan *data mining* algoritma C4.5 dalam menyeleksi data, mencari pola menarik dan mengevaluasi data yang menghasilkan suatu keputusan penentuan kelayakan anggota *security*.
2. Untuk mendeskripsikan hasil analisis data dari *software RapidMiner* dan keakuratan pohon keputusan yang dihasilkan dari hasil perhitungan.

1.6 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi setiap pembaca. Adapun manfaat dari penelitian ini sebagai berikut:

1.6.1 Manfaat Teoritis

Hasil penelitian ini diharapkan dapat digunakan sebagai bahan referensi, masukan, perbandingan dalam melakukan penelitian yang lebih dalam dan luas atau bahan pertimbangan dalam penelitian yang selanjutnya serta menambah wacana keilmuan di bidang teknik informatika.

1.6.2 Manfaat Praktis

Selain manfaat teoritis, penelitian ini juga diharapkan dapat membawakan manfaat praktis seperti dibawah ini:

1. Peneliti

Hasil penelitian ini diharapkan bermanfaat dan memberikan kontribusi keilmuan pada penelitian bidang klasifikasi *data mining*.

2. Manajemen

Hasil penelitian ini diharapkan dapat membantu direktur dan manajer PT Avava Duta Indonesia untuk memperoleh suatu pengetahuan baru dalam pengambilan keputusan penentuan anggota *security* yang layak mengingat anggota *security* mencerminkan keunggulan jasa dari perusahaan.

BAB II KAJIAN PUSTAKA

2.1 *Knowledge Discovery in Database (KDD)*

Pemrosesan data secara komputerisasi telah diimplementasikan 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*.

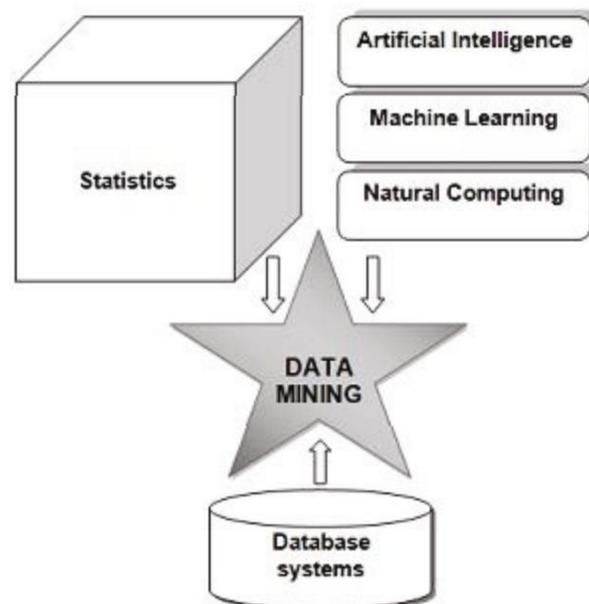
Data mining merupakan suatu solusi yang dapat menemukan pola atau pengetahuan yang bermanfaat secara otomatis dari sekumpulan data yang berjumlah banyak (Sunjana, 2010). Untuk istilah *data mining* kadang disebut juga dengan *Knowledge Discovery in Databases (KDD)*. *Knowledge Discovery in Databases (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.

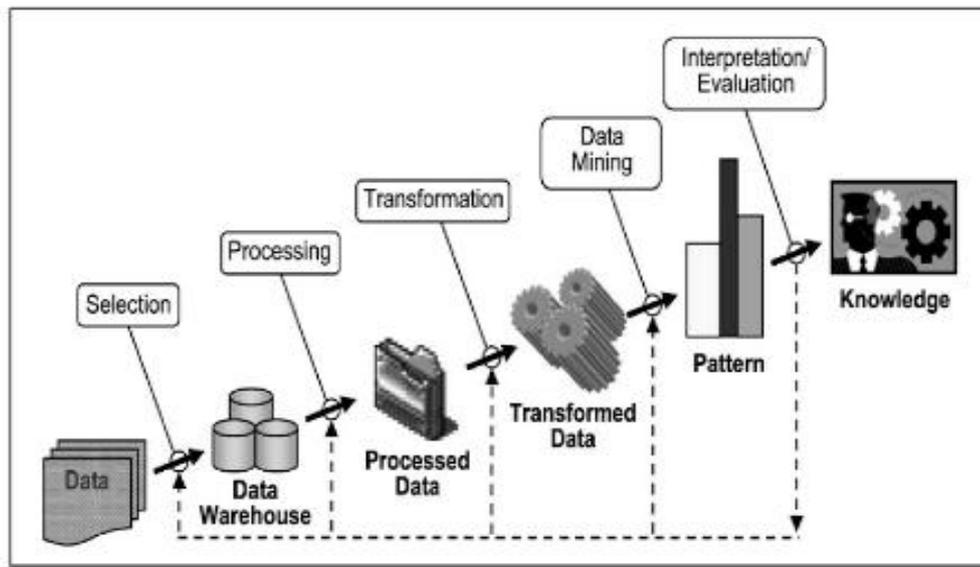
Menurut Larose (2005), terdapat beberapa faktor yang mempengaruhi kemajuan dalam bidang *data mining* antara lain (Syarif, 2015):

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 perubahan data sebelum diproses dalam *data mining*. Perubahan 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*

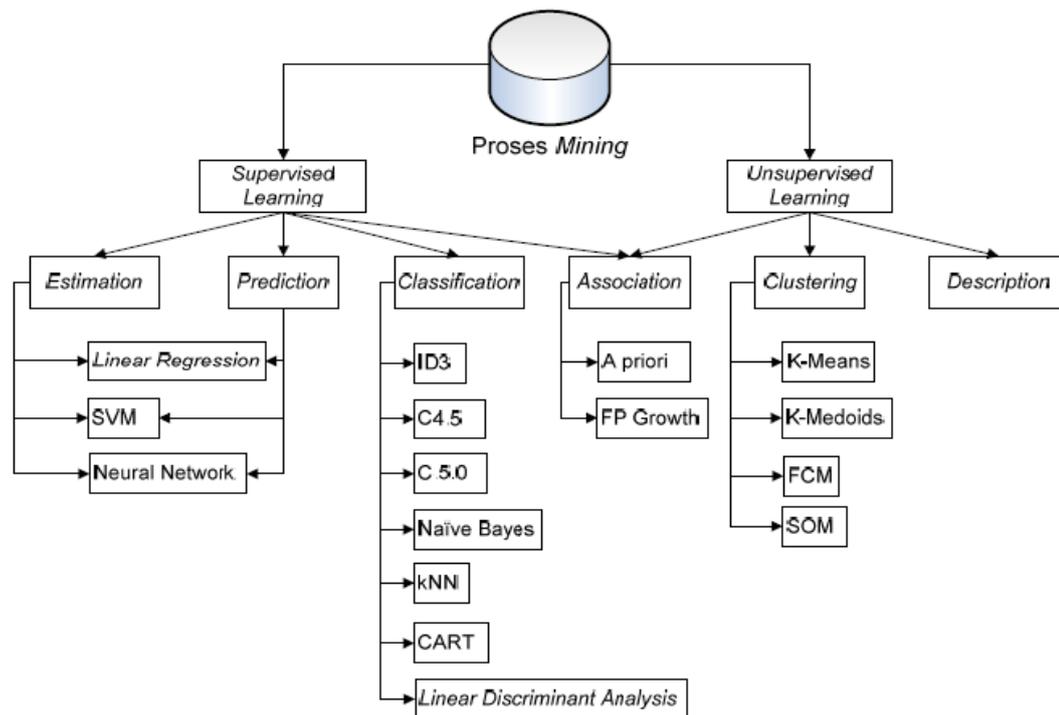
Data mining dapat digunakan untuk melihat hubungan antara penjualan satu produk dengan produk lainnya. Dengan adanya penerapan *data mining* dalam segi *cross market analysis* maka dapat dioptimalkan penjualan produk.

d. Profil *Customer*

Data mining dapat digunakan untuk memantau profil *customer* 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 pengelompokan 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 atribut 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:

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

Rumus 2.1 Perhitungan Entropi

Keterangan:

S : himpunan kasus

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

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

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

Rumus 2.2
Perhitungan *Gain*

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

$|S_i|$: 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:

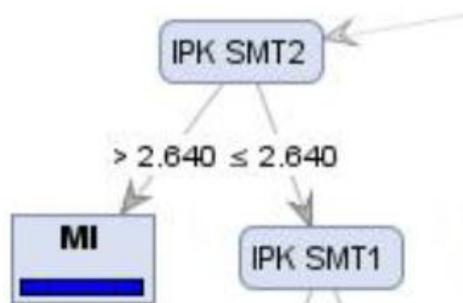
$$GainRatio(X) = \frac{Gain(X)}{SplitInfo(X)}$$

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 menunjukkan 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. Pengelompokan 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 algoritma C4.5 dapat digunakan pada *dataset* pelanggan ke dalam

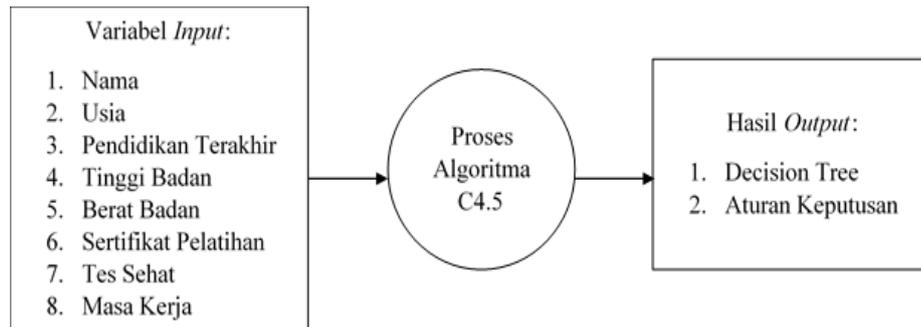
kegiatan manajemen strategi sehingga dapat mempertahankan loyalitas pelanggannya.

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 DEHASSEN 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:



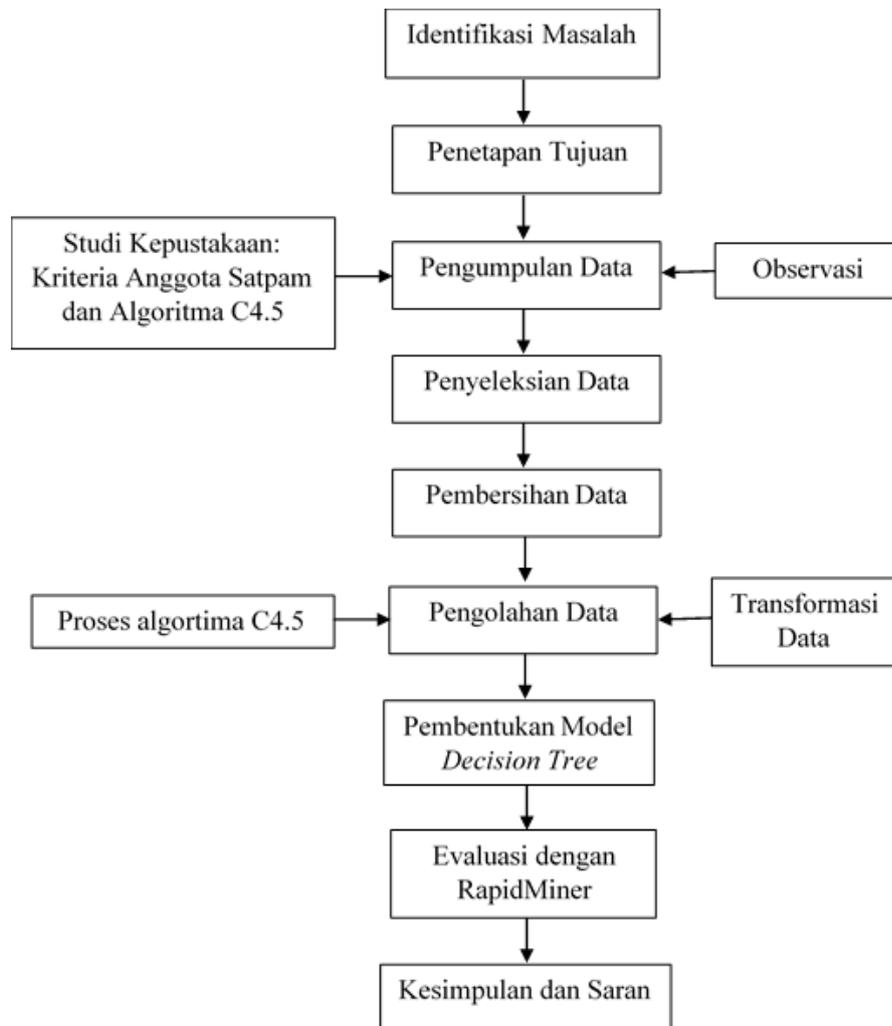
Gambar 2.5 Kerangka Pemikiran

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*.

BAB III METODE PENELITIAN

3.1 Desain Penelitian

Dalam melakukan penelitian, dibutuhkan desain penelitian agar penelitian yang dilakukan dapat berjalan dengan baik. Berikut ini merupakan desain penelitian yang digunakan:



Gambar 3.1 Desain Penelitian

Adapun penjelasan dari desain penelitian diatas meliputi tahapan-tahapan berikut ini:

1. Identifikasi masalah

Menjelaskan hal-hal yang menjadi permasalahan bagi peneliti untuk melakukan penelitian untuk menyelesaikan masalah tersebut.

2. Penetapan tujuan

Dari permasalahan yang ada tersebut maka penulis menetapkan tujuan penelitiannya. Tujuan penelitian berisi hal akan akan dicapai dari penelitian ini.

3. Pengumpulan data

Data primer akan dikumpulkan oleh peneliti untuk mencapai hasil penelitian. Data dapat dikumpulkan melalui tahap observasi langsung ke instansi jasa pelayanan *security*. Untuk data sekunder diperoleh dengan menggunakan referensi dari jurnal dan buku yang berkaitan dengan kriteria anggota *security* dan algoritma C4.5 sebagai studi kepustakaan.

4. Penyeleksian data

Proses pemilihan data yang dilakukan adalah dengan memilih data penelitian yang akan diproses dalam proses *data mining*. Pada proses ini dipilih mana data sampel yang mewakili populasi yang ada.

5. Pembersihan data

Setelah data dipilih, kemudian proses pembersihan data untuk menghilangkan duplikasi data, menghilangkan kesalahan pada data seperti salah cetak (tipografi).

6. Pengolahan data

Pengolahan data meliputi transformasi data yaitu mengubah bentuk data penelitian yang masih berupa angka (kuantitatif) menjadi data berupa kategori (kualitatif). Dari data yang telah dikumpulkan maka dilakukan perhitungan manual dengan menggunakan proses algoritma C4.5 untuk menentukan kelayakan personel *security*. Perhitungan tersebut meliputi pencarian nilai entropi, *gain*, *split info*, dan *gain ratio*.

7. Pembentukan model *Decision Tree*

Berdasarkan perhitungan yang telah dilakukan maka dibentuk sebuah pohon keputusan, pembentukan pohon dengan menentukan terlebih dahulu simpul terpilih, yaitu dengan menghitung nilai *information gain* dari masing-masing atribut. Simpul yang terpilih nantinya dibagi menjadi simpul akar, simpul internal, dan simpul daun.

8. Evaluasi dengan *RapidMiner*

Hasil perhitungan yang diperoleh dari pengolahan secara manual akan dimasukkan ke dalam *software RapidMiner* untuk dievaluasi dan dianalisis tingkat akurasi.

9. Kesimpulan dan Saran

Penarikan kesimpulan merupakan tahap akhir dari penelitian, dari keseluruhan proses penelitian nantinya dapat diketahui cocok atau tidak cocoknya penerapan *data mining* dalam penentuan kelayakan anggota *security*.

3.2 Teknik Pengumpulan Data

Teknik-teknik yang digunakan oleh peneliti dalam pengumpulan data adalah sebagai berikut:

1. Kajian pustaka

Pada penelitian mengenai penentuan kelayakan personil *security* menggunakan algoritma C4.5, kajian pustaka yang dilakukan adalah mengumpulkan data-data dari berbagai media seperti buku bacaan, jurnal, maupun artikel-artikel yang berasal dari internet yang berkaitan dengan metode algoritma C4.5. Adapun sasaran dari metode kajian pustaka ini adalah memahami algoritma C4.5 dan juga memahami sistematika penentuan kriteria seorang anggota *security* yang layak.

2. Observasi

Metode observasi atau metode studi lapangan adalah metode pengamatan langsung ke lokasi penelitian. Metode ini bertujuan untuk mengumpulkan data-data anggota *security* dan calon-calon *security* di PT Avava Duta Indonesia.

3.3 Operasional Variabel

Variabel yang digunakan pada penelitian dibedakan menjadi variabel bebas dan variabel terikat. Dari data yang dikumpulkan, data yang diambil sebagai atribut atau variabel terikat adalah atribut kelayakan anggota *security*. Atribut-atribut yang digunakan sebagai variabel bebas ada beberapa diantaranya seperti atribut

nama, usia, pendidikan terakhir, tinggi badan, berat badan, tes sehat, sertifikat pelatihan dan masa kerja. Penjelasan atribut-atribut yang berperan sebagai variabel bebas yaitu:

1. Nama, berfungsi sebagai variabel ID dengan tipe *text* yang membedakan variabel yang satu dengan yang lain.
2. Usia, berfungsi sebagai variabel atribut dengan tipe *polynomial* yang berada pada *range* 19 tahun sampai dengan 54 tahun.
3. Pendidikan Terakhir, terdiri atas status pendidikan SD, SMP, SMA dan Sarjana, yang berfungsi sebagai variabel atribut dengan tipe *binominal*.
4. Tinggi Badan, berfungsi sebagai variabel atribut dengan tipe *binominal* yang disesuaikan dengan standar tinggi badan anggota *security*.
5. Berat Badan, variabel ini disesuaikan nilainya terhadap variabel tinggi badan yang berfungsi sebagai variabel atribut dengan tipe *polynomial* dengan memperhatikan standar BMI (*Body Mass Index*).
6. Tes Sehat, berfungsi sebagai variabel atribut dengan tipe *binominal* yang berfungsi untuk menjelaskan fisik anggota.
7. Sertifikat Pelatihan, terdiri atas status pelatihan gada pratama, gada madya dan gada utama yang berperan sebagai variabel atribut dengan tipe *binominal*.
8. Masa Kerja, berfungsi sebagai variabel atribut dengan tipe *polynomial* yang menandakan lama kerja dan pengalaman dari personil tersebut.

Data-data yang akan diambil dalam penelitian yang siap untuk proses klasifikasi dengan atribut dan nilai atribut yang ada dapat dilihat pada tabel 3.1:

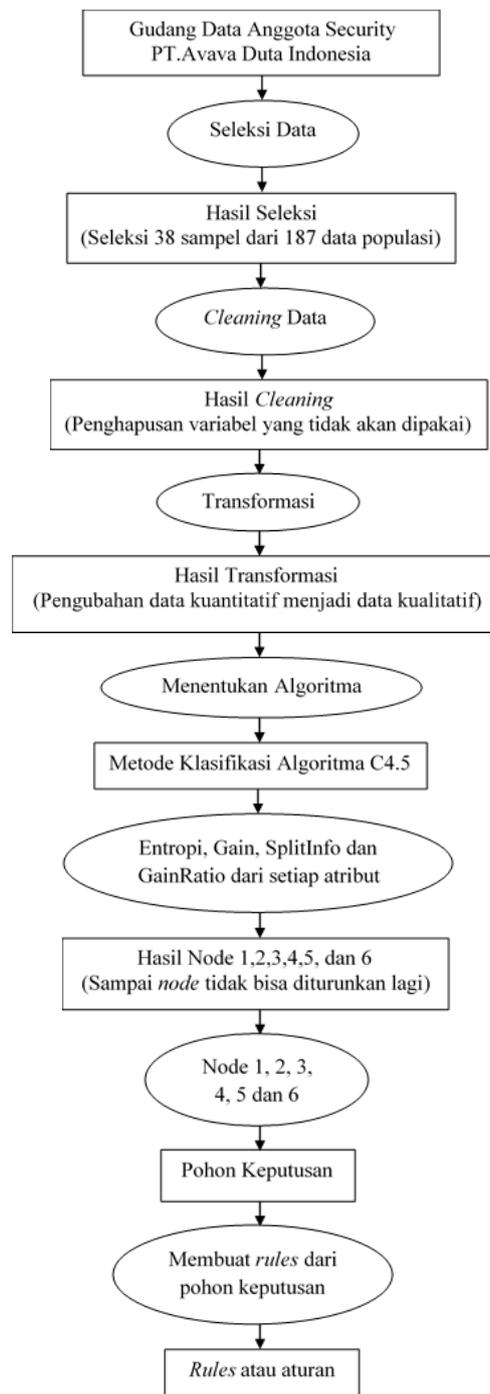
Tabel 3.1 Data Karyawan PT Avava Duta Indonesia

No	Nama	Tempat & Tanggal Lahir	Usia	JK (P/L)	Pendidikan Terakhir	Tinggi Badan	Berat Badan	Tes Sehat	Sertifikat Pelatihan	Masa Kerja
1	Megi DaryantoP	Tapan, 02 Nov 1985	33	L	SMA	179	86	Sehat	-	1 bulan
2	Syukurman T	Hilina'a, 06 Okt 1993	25	L	SMA	165	65	Sehat	Gada Pratama	1 tahun 10 bulan
3	Zulkarnain Lain	Air Buaya, 10 Nov 1989	29	L	SMA	163	60	Sehat	-	3 tahun 10 bulan
4	Syaiful Asnadi	Kerinci, 01 Des 1968	50	L	Sarjana	173	75	Sehat	Gada Madya	3 tahun
5	Refi Satrianto	Pasar Usang, 31 Okt 1990	28	L	SMA	177	83	Sehat	Gada Pratama	5 tahun
6	Mardianto	Medan, 07 Jan 1983	35	L	SMA	170	80	Sehat	Gada Pratama	3 tahun
7	Ramlan	Padang, 28 Okt 1981	37	L	SMA	171	82	Sehat	Gada Pratama	2 tahun
8	Ardiansyah	Pantai Harapan, 08 Apr 1989	29	L	SMA	177	86	Sehat	Gada Pratama	1 tahun
9	Edi Sud	Concong Luar, 11 Jun 1984	34	L	SMK	170	90	Sehat	Gada Pratama	3 tahun 8 bulan
10	Yoga Sugama	Palembang, 21 Agus 1989	29	L	SMA	175	82	Sehat	-	2 tahun
11	Brando	P. Sidempuan, 17 Sept 1978	40	L	SMA	168	60	Sehat	Gada Pratama	4 tahun 2 bulan
12	Rio Reno	Jambi, 21 Mar 1987	31	L	SMA	173	72	Sehat	Gada Pratama	2 tahun
13	Yuliadin	Semarang, 31 Des 1988	30	L	SMA	180	80	Sehat	Gada Pratama	2 tahun
14	Samdani	Aceh, 01 Feb 1991	27	L	SMA	177	70	Sehat	Gada Pratama	1 tahun
15	Abdul	Manado, 01 Feb 1993	25	L	SMA	175	72	Sehat	Gada Pratama	1 tahun
16	Alek	Palembang, 01 Sept 1995	23	L	SMA	170	72	Sehat	-	1 tahun
17	Ganti Hasibuan	Medan, 03 Feb 1991	27	L	SMA	179	84	Sehat	-	1 tahun
18	Wilhimus	Flores, 01 Feb 1990	28	L	SMA	176	68	Sehat	-	1 tahun
19	Sirajuddin	Makasar, 20 Feb 1985	33	L	SMA	168	72	Sehat	-	4 tahun
20	Ramadhani A	Sidomulyo, 14 Mei 1986	32	L	SMA	171	66	Sehat	-	3 tahun
21	Jhon Hendri	Taklawi, 04 Jul 1977	41	L	SMA	169	66	Sehat	Gada Pratama	1 tahun
22	Robbi Saputra	Tanjung Pinang, 05 Jun 1989	29	L	SMA	175	80	Sehat	Gada Pratama	3 tahun

Sumber: Data olahan penulis, 2017

3.4 Metode Analisis dan Rancangan Sistem

Berikut adalah gambaran rancangan sistem pengolahan *data mining*:



Gambar 3.2 Rancangan Sistem

Dari gambar 3.2 diketahui bahwa proses akan dimulai dengan penyeleksian data pada gudang data. Setelah memperoleh data seleksi sebanyak 38 sampel data dari total 187 data populasi, dilanjutkan dengan proses *cleaning* data untuk menghapus variabel yang tidak digunakan. Selanjutnya adalah proses transformasi data ke dalam bentuk yang dapat diolah. Dari hasil transformasi tersebut maka diproses dengan menentukan algoritma yang akan digunakan.

Metode analisis yang digunakan dalam penelitian adalah teknik klasifikasi dengan algoritma C4.5, karena dalam klasifikasi terdapat *target* variabel kategori. Hasil transformasi tersebut dihitung dengan menggunakan algoritma C4.5 meliputi pencarian nilai entropi, *gain*, *split info* dan *gain ratio* untuk membentuk suatu pohon keputusan. Proses akan dilakukan berulang kali hingga *node* pada pohon keputusan tersebut tidak dapat diturunkan lagi. Dari pohon keputusan tersebut akan dibuat aturan keputusannya. Sistem yang akan dirancang digunakan untuk menentukan kelayakan dari seorang personil *security*.

3.5 Lokasi dan Jadwal Penelitian

Adapun lokasi penelitian dan jadwal penelitian akan dijelaskan sebagai berikut:

3.5.1 Lokasi Penelitian

Penelitian akan dilakukan di PT Avava Duta Indonesia yang berlokasi di Mall Jodoh Marina Blk B 421-422, Jodoh, Batam.

3.5.2 Jadwal Penelitian

Penelitian ini dimulai dari September 2017 sampai dengan Januari 2018, dengan rincian jadwal penelitian sebagai berikut:

Tabel 3.2 Jadwal Penelitian

No	Kegiatan	Sept 2017	Oktober 2017	November 2017	Desember 2017	Januari 2018	
1	Bimbingan	█					
2	Studi Kepustakaan	█	█	█	█	█	
3	Penentuan Judul	█	█	█	█	█	
4	Pengajuan Proposal	█	█	█	█	█	
5	Pengumpulan Data	█	█	█	█	█	
6	Pengolahan Data	█	█	█	█	█	
7	Penyusunan Laporan Penelitian	█	█	█	█	█	

Sumber: Data olahan penulis, 2017