

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori**

Peneliti menguraikan beberapa teori atau prinsip dasar yang dipakai guna mendukung pengambilan hasil, adapun teori yang dimaksud terdiri atas dua teori besar yaitu teori umum dan teori khusus, adapun uraiannya, antara lain:

##### **2.1.1 Teori Umum**

Sebuah penelitian harus didasarkan pada teori-teori yang telah dikemukakan sebelumnya, salah satunya ialah teori umum. Teori umum ini dimaksudkan untuk memperkuat isi dari teori yang ada hal ini dimaksudkan supaya penelitian yang dilakukan memperoleh hasil yang lebih baik. Adapun isi dari teori umum tersebut dapat dijelaskan sebagai berikut:

##### **2.1.1.1 *Knowledge Discovery in Database (KDD)***

*Knowledge Discovery In Database (KDD)* diartikan sebagai sistem guna menemukan informasi dari *database* yang tersedia. *Database* berisi tabel-tabel yang berkolerasi. Informasi-informasi yang telah ditemukan pada *database* tersebut dikonversikan menjadi sebuah basis pengetahuan (*knowledge base*) guna dijadikan sebagai pertimbangan dalam mengambil keputusan.

*Knowledge Discovery in Database (KDD)* dan *data mining* dikolaborasikan oleh pihak-pihak berkepentingan sesuai keperluan pada kondisi tertentu guna menjabarkan operasi penambangan informasi yang potensial tetapi

belum ditemui pada sebuah basis data yang cukup besar. Dua komponen tersebut mempunyai konsepsi yang tidak sama, namun memiliki kolerasi diantara keduanya, bahkan salah satu hierarki dari prosedur KDD ialah *data mining*. Berikut prosedur KDD yang dijabarkan seperti dibawah ini:

1. *Data Selection*

Tahap awal yang perlu dijalankan ialah melakukan seleksi pada tahapan *Knowledge Discovery in Database*. Setelah diseleksi, hasil yang didapatkan dilakukan penyimpanan secara terpisah dari basis data operasional yang akan dipergunakan sebagai bahan pendukung *data mining*.

2. *Pre-processing / Cleaning*

Tahap selanjutnya merupakan tahap yang perlu diperhatikan, yakni proses *cleaning*. Hal-hal yang harus dilakukan pada proses *cleaning* meliputi membuang data yang terdupliaksi, melakukan pemeriksaan pada data yang tidak konsisten, serta melakukan perbaikan pada data yang dianggap salah atau tidak sesuai, contohnya ialah kesalahan pencetakan. Selanjutnya, terdapat sebuah proses yang bernama *enrichment*, yakni tahapan dimana “memperkaya” data menggunakan data lain yang sejalan dengan data yang telah tersedia sebelumnya guna digunakan pada *Knowledge Discovery in Database* (KDD), semisal informasi eksternal.

3. *Transformation*

Pada tahapan ini, terdapat proses kunci yang biasa disebut *coding* sehingga data yang telah ada sesuai dengan keperluan pada saat pemrosesan *data mining*. Proses ini meliputi mengkategorian data sehingga membentuk pola-pola yang memudahkan pengambilan informasi berdasarkan jenis data dan tergantung apa yang dikehendaki oleh pengolah data.

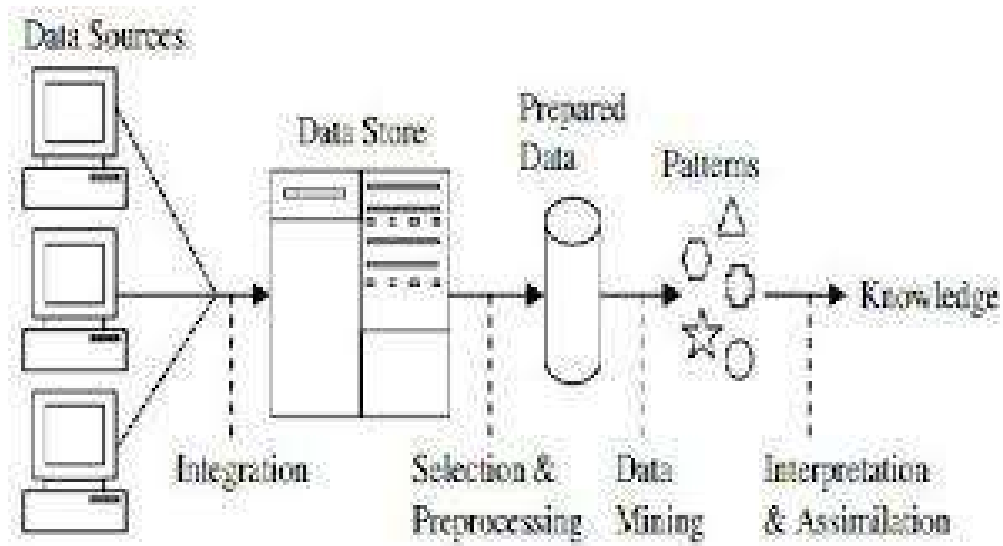
#### 4. *Data Mining*

*Data mining* dapat diartikan sebagai pencarian sebuah corak tertentu atas data yang telah dipilih sebelumnya melalui sistem atau kaidah yang ada. Terdapat varian cara atau algoritma yang dapat dipergunakan pada *Data Mining*. Cara dan algoritma yang ada dipilih sesuai dengan sasaran dan keperluan pada *Knowledge Discovery in Database (KDD)* yang akan dilakukan oleh pengolah data.

#### 5. *Interpretation / Evaluation*

Corak yang terbentuk pada tahap transformasi dan *data mining* harus menghasilkan sebuah hasil yang memudahkan pengambilan keputusan dari semua pihak yang memiliki kepentingan. Oleh karena itu, proses ini adalah proses terakhir yakni *interpretation*. Pada tahapan ini, dilakukan evaluasi pada informasi serta pola apakah data berkontradiktif dengan yang sebenarnya terjadi atau jawaban sementara yang telah dimuskani.

Pendapat dari (Mardi, 2016) proses KDD atau *Knowledge Discovery in Database* mampu dibuat skemanya sebagai berikut:



**Gambar 2.1** Proses Dalam KDD atau *Knowledge Discovery in Database*

**Sumber:** (Mardi, 2016)

### 2.1.1.2 Data Mining

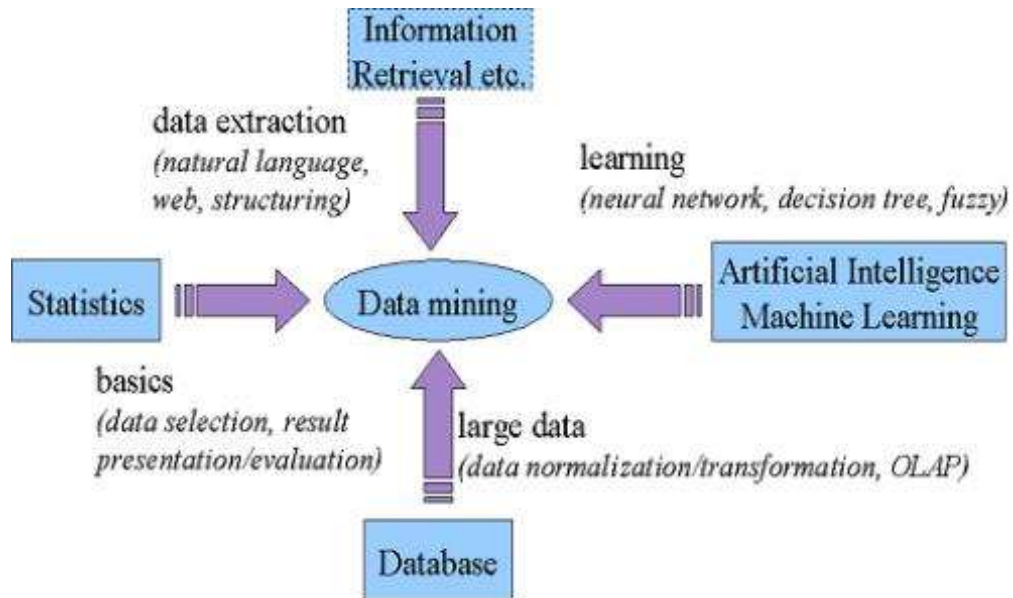
Seperti yang dikemukakan oleh Gartner Group, *data mining* dapat diartikan sebagai tahapan yang dilakukan guna mencari kolerasi yang memiliki pengertian, corak dan kualifikasi tertentu dengan pemilihan dari begitu banyak data kemudian dikumpulkan pada sebuah media penyimpanan berbasis teknologi statistik atau arimatika. *Data mining* sendiri merupakan kolaborasi dari berbagai pengetahuan yakni machine learning, pengenalan pola, statistik, database, dan visualisasi guna menangani masalah-masalah yang terjadi pada pengambilan data seharusnya mampu digunakan untuk pengambilan keputusan.

Berdasarkan pendapat dari David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT, menyatakan bahwa *data mining* dapat dikatakan sebagai proses penganalisisan terhadap data yang berskala besar guna mencari kolerasi antara

data tersebut dengan membuat kesimpulan atas data tersembunyi sehingga ditemukan dan mampu dipakai oleh pihak yang berkepentingan. *Data mining* biasanya terdiri dari beberapa komponen yakni statistik, aritmatika, *artificial intelligence*, dan keilmuan tentang mesin guna memilah-milah dan melakukan identifikasi data sesuai dengan keilmuan yang dibutuhkan. *Data mining* dikatakan sebagai berbagai rangkaian tahapan guna menelusuri informasi-informasi baru berbentuk keilmuan apabila tahapan dilakukan secara manual. Terdapat unsur-unsur yang berkaitan dengan *data mining*, yakni sebagai berikut:

1. *Data mining* dapat dikatakan sebagai data yang transformasikan secara otomatis.
2. Data yang dilakukan pemrosesannya biasanya berskala besar.
3. Tujuan *data mining* ialah memperoleh kolerasi dengan motif tertentu sehingga ditemukan informasi yang mampu digunakan untuk pengambilan keputusan.

*Data mining* sudah sering digunakan tetapi tidak terlalu terkenal. Kelemahan dari *data mining* ialah membutuhkan banyak keilmuan dari sektor-sektor yang cukup banyak sebagai acuan atau dasar pengolahannya serta wajib terbukti kebenarannya. Gambar dibawah ini menjelaskan bahwa *data mining* memiliki berakar dari banyaknya aspek keilmuan meliputi kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik, *database*, dan juga *information retrieval*.



**Gambar 2.2** Bidang Ilmu Dari Data Mining

**Sumber:** (Mardi, 2016)

### 2.1.1.3 Pengelompokan *Data Mining*

Dilihat dari peran dan fungsinya, *Data mining* dapat dikelompokkan kedalam tugas berikut:

1. *Description* (Deskripsi)

Analisis melakukan upaya-upaya guna menemukan teknik sederhana atau yang memudahkan sehingga menghasilkan penggambaran corak atau motif tertentu dan tren yang terjadi pada data yang diteliti. Contohnya ialah pada pemilihan umum dalam memilih kepala negara, yakni kemungkinan-kemungkinan yang terjadi atas fakta atau penjelasan ilmiah mengenai seperti apa pilihan pemilih. Pendeskripsian atas motif dan kecenderungan memberikan penjelasan-penjelasan tertentu mengenai hal tersebut.

## 2. *Estimation* (Estimasi)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pasca sarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.

## 3. *Prediction* (Prediksi)

*Prediction* tidak jauh berbeda dengan dugaan-dugaan dan kemungkinan yang ada, terkecuali adanya jawaban sementara atas apa yang terjadi dikemudian hari. Dapat diberikan contoh sebagai berikut:

- a. Estimasi atas harga beras dalam dua bulan kedepan .
- b. Estimasi tingkat kemiskinan dalam 5 tahun kedepan.
- c. Estimasi angka harapan hidup dalam beberapa tahun kedepan apabila ketersediaan akan pangan mengalami peningkatan.

Terdapat berbagai sistem dan cara-cara yang dimampu diterapkan terhadap menghasilkan dugaan dan kemungkinan untuk diprediksi.

## 4. *Classification* (Klasifikasi)

Pada fungsi ini, analisis ditemukan sasaran variabel berkategori. Semisal penggolongan pendapatan dikategorikan kedalam tiga tingkatan, yakni

pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Terdapat contoh lainnya, yakni:

- a. Mendiagnosis penggunaan dengan kartu debit merupakan hal yang efektif atau tidak.
- b. Mendiagnosis pengajuan hutang kepada bank dalam jumlah besar mampu memberikan keuntungan atau kerugian pada perusahaan.
- c. Mendiagnosis kadar obat yang cocok sesuai dengan penyakit yang diderita oleh pasien.

#### 5. *Clustering* (Pengkusteran)

Pengkusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain.

Pengkusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengkusteran. Pengkusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengkusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal. Contoh pengkusteran dalam bisnis dan penelitian adalah :

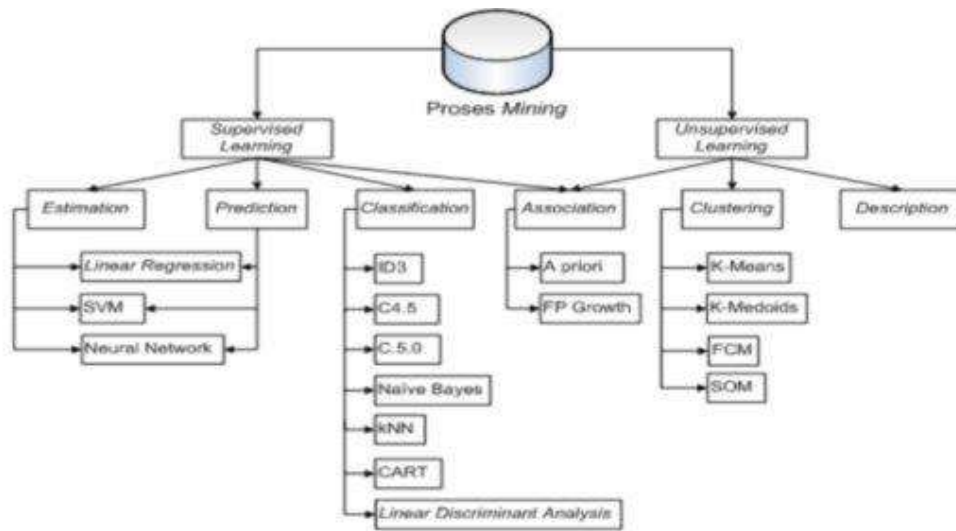


- a. Mengelompokkan target pasar kemudian konsumen-konsumen potensial sehingga mampu menghemat biaya pemasaran.
- b. Mengelompokkan beberapa transaksi yang tidak wajar pada kegiatan operasional perusahaan dan biasanya dilakukan oleh akuntan sebuah perusahaan.
- c. Mengelompokkan kriteria-kriteria dari gen akan penyakit bawaan orangtua yang akan diturunkan kepada anaknya.

6. *Association* (Asosiasi)

*Association* memiliki peran untuk mendapatkan instrumen-instrumen yang berasosiasi satu dengan yang lainnya pada waktu tertentu. Istilah lain yang digunakan ialah analisis keranjang belanja. Berikut contoh-contoh *association* :

- a. Melakukan penelitian dari pengguna akses internet sehingga mampu menerima tanggapan positif ketika dilakukan promosi-promosi dan peningkatan paket internet yang digunakan.
- b. Melakukan pengamatan atas barang-barang kosmetik yang memiliki ketertakaitan satu dengan yang lainnya sehingga memungkinkan barang tersebut dibeli dalam suatu waktu yang sama.



**Gambar 2.3** Pengelompokan Teknik Data Mining

**Sumber:** (Hermawan et al., 2019)

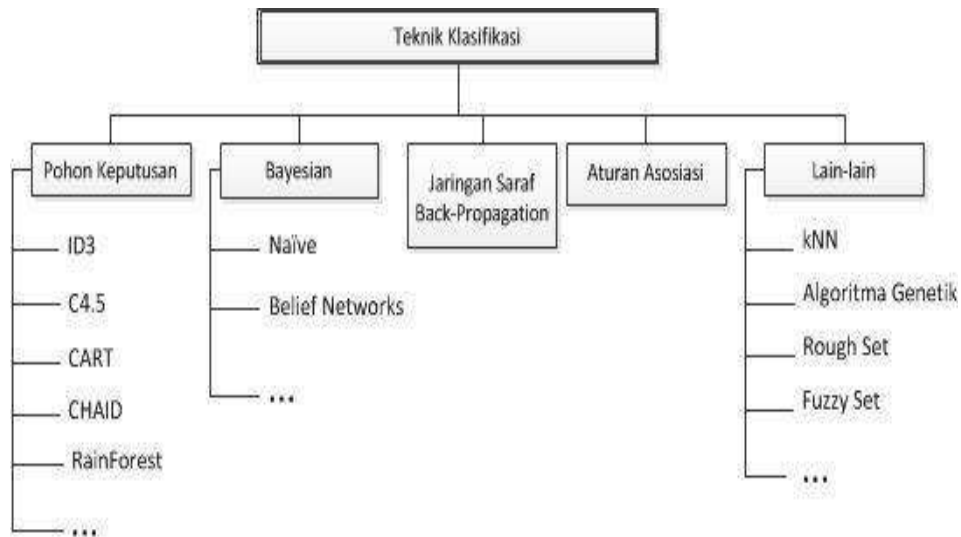
#### 2.1.1.4 Klasifikasi

Seperti yang dikemukakan oleh (Azwanti, 2018), klasifikasi data dapat dikatakan sebagai kegiatan menyatukan instrumen-instrumen yang sama pada suatu kelompok atas objek-objek yang diambil pada suatu basis data dan mengkategorikannya kedalam golongan-golongan tertentu seperti yang dikehendaki oleh analis. Klasifikasi bertujuan guna menentukan model dari training set yang memilah-milah atribut ke dalam kelas yang sesuai, kemudian atribut tersebut diklasifikasikan ke kelas yang belum ada sebelumnya. Terdapat metode-metode pada klasifikasi umum, yakni pengklasifikasi pohon keputusan, pengklasifikasi *Bayesian*, pengklasifikasi *k-nearest neighbour*, penalaran berbasis kasus, algoritma genetika dan teknik logika *fuzzy*.

Satu dari banyaknya tujuan penggunaan *data mining* memanglah klasifikasi. Pengklasifikasian dilakukan untuk kali pertama pada bidang biologi yakni mengklasifikasikan spesies makhluk hidup, pencetusnya ialah Carolus von Linne (atau dikenal dengan nama Carolus Linnaeus) yakni melakukan klasifikasi spesies hewan berdasarkan karakteristik fisiknya sehingga sampai hari ini, ia dikenal sebagai Bapak Klasifikasi.

Berikut model-model yang mendukung proses pengklasifikasian data, yakni sebagai berikut:

- a. Pohon keputusan
- b. Pengklasifikasi *bayes/naive bayes*
- c. Jaringan saraf tiruan
- d. Analisis statistik
- e. Algoritma genetik
- f. *Rough sets*
- g. Pengklasifikasi *k-nearest neighbour*
- h. Metode berbasis aturan
- i. *Memory based reasoning*
- j. *Support vector machine*



**Gambar 2.4** Pengelompokan Teknik Klasifikasi

**Sumber:** (Della, 2017)

### 2.1.1.5 Pohon Keputusan

Salah satu bentuk metode dalam pengklasifikasian data ialah metode pohon keputusan atau *decission tree*. Metode ini dapat diartikan sebagai sebuah metode yang mengubah data dilapangan dalam jumlah besar menjadi sebuah pohon keputusan yang mencerminkan sebuah pilihan-pilihan keputusan. [iluhan-piluhan keputusan tersebut memudahkan para pemegang kepentingan memahami data yang menjadi bahasa alami.

Pohon keputusan itu sendiri digunakan untuk melakukan pembagian pada banyaknya data yang ada menjadi kelompok-kelompok kecil secara terstruktur sehingga menghasilkan pilihan-pilihan guna mendapatkan sebuah kesimpulan. Dikarenakan adanya pembagian anggota-anggota kelompok maka hasil yang

ditunjukkan menjadi serupa. Pohon keputusan biasanya diterjemahkan ke dalam bentuk tabulasi dengan atribut dan record. Atribut sendiri diartikan sebagai suatu acuan sebagai satuan pengkategorian yang membentuk pohon keputusan itu sendiri. Contohnya ialah seseorang yang ingin bermain sepak bola harus memperhatikan kriteria-kriteria tertentu seperti cuaca, kondisi lapangan, dan lainnya. Terdapat atribut yang memberikan solusi dari masing-masing butir data yakni dinamakan target atribut. Nilai yang terkandung di dalam atribut disebut instance. Contohnya ialah atribut kondisi lapangan mempunyai instance yakni lapangan yang rata, berlubang, berlumpur, dan lain sebagainya. Secara garis besar, peran pohon keputusan ialah mentransformasikan data yang ada menjadi tabel, tabel disusun menjadi sebuah model pohon, model pohon yang ada diklasifikasi menjadi *rule*, dan terakhir membuat *rule* menjadi suatu bentuk yang lebih sederhana dan mudah dimengerti.

Terdapat beberapa algoritma yang biasanya membentuk sebuah pohon keputusan, yakni ID3, CART, dan C4.5. Algoritma C4.5 merupakan turunan dari algoritma ID yang dikembangkan lebih lanjut. Sebuah pohon keputusan dibentuk seperti sebuah pohon yang asli dengan bagian-bagian sebagai berikut:

1. *Root Node* : berada paling atas pada sebuah pohon keputusan.
2. *Internal Node* : seperti cabang yang bersumber pada satu input dan mengeluarkan dua output lagi.
3. *Leaf Node* : berada pada ujung pohon yang memiliki satu input dan tidak bercabang lagi

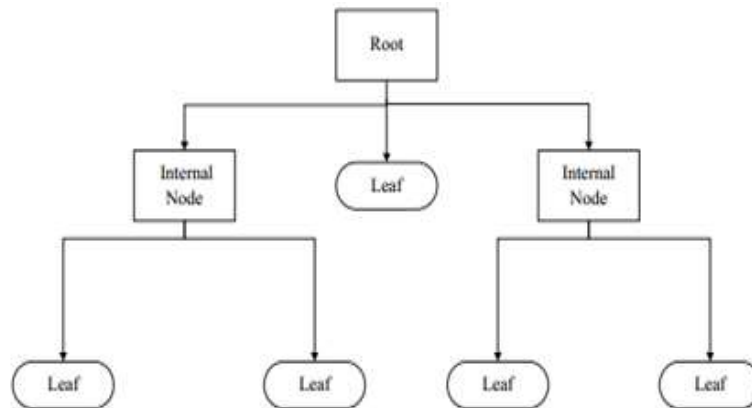
Seperti yang dijelaskan oleh Triisant, terdapat beberapa kelebihan dan kekurangan pada penggunaan pohon keputusan ini, antara lain:

1. Kelebihan Pohon Keputusan:

- Kekompleksan area yang digunakan untuk mengambil keputusan dapat dibuat menjadi lebih sederhana.
- Menyederhanakan berbagai perhitungan data dikarenakan terjadinya eliminasi pada data yang tidak diperlukan oleh analisis.
- Proses pemilihan fitur dari *internal node* yang berbeda lebih fleksibel.
- Meminimalisir kesalahan yang mungkin memunculkan masalah baru karena pohon keputusan menyederhanakan kriteria yang ada tanpa mengurangi mutu hasil yang akan didapatkan.

2. Kekurangan Pohon Keputusan :

- Adanya kemungkinan terjadinya *overlap* yang disebabkan banyaknya hasil keputusan dan kriteria dalam jumlah yang relatif besar. Melalui kejadian ini akan memakan waktu untuk pengambilan keputusan lebih lama dan membutuhkan kapasitas penyimpanan yang semakin besar.
- Adanya jumlah *error* yang banyak pada setiap pohon keputusan.
- Adanya kesulitan dalam merancang pohon keputusan yang menunjukkan hasil maksimal.
- Mutu keputusan yang diambil sangat bergantung pada desain pohon keputusan tersebut dibuat.



**Gambar 2.5** Model Pohon Keputusan

**Sumber:** (Hamidah, 2012)

#### 2.1.1.6 Algoritma C4.5

Algoritma C4.5 diaplikasikan guna mendukung pohon keputusan. (Bening, 2014) mengemukakan sebuah pendapat bahwa algoritma C4.5 dikategorikan sebagai satu dari banyaknya *machine learning*. Mesin atau dalam hal ini komputer yang menggunakan algoritma C4.5 menunjukkan hasil kelompok data yang mampu dipelajari dengan nama *learning dataset*. Algoritma C4.5 menghasilkan sebuah kelompok data yang terbagi kedalam kelas-kelas tertentu melalui proses *test dataset*.

Hasil dari pengolahan yang telah dijabarkan pada bagian sebelumnya akan diolah kembali dengan sebutan *test dataset*. Algoritma C4.5 menjadi acuan dasar guna membuat pohon keputusan yang dinagi enjadi atribut-atribut yang berprioritas tertinggi atau dapat dikatakan mempunyai nilai *gain* tertinggi yang diambil dari nilai *entropy atribut* sebagai pedoman dari pengklasifikasian atribut yang telah ditetapkan sebelumnya. Setelah itu, dibentuklah cabang-cabang pohon

yang dilakukan perluasan sehingga menyerupai seperti pohon. Tertulis dalam kamus IGI Global (*International Publisher of Progressive Academic*), *entropy* ialah banyaknya data-data yang tidak sejalan atau relevan dengan data yang telah pada suatu himpunan data. Gain dapat didefinisikan sebagai data yang diambil yang dipengaruhi oleh adanya perubahan *entropy* pada himpunan data, dapat karena adanya observasi tertentu atau adanya keikutsertaan data tersebut pada suatu set data.

Seperti yang dikemukakan oleh (Andriani, 2012) menyatakan bahwa terdapat empat tahapan C4.5 yang membentuk suatu pohon keputusan, antara lain:

1. Mempersiapkan *data training*. *Data training* dapat diperoleh dari data-data yang sudah ada sebelumnya (*data history* atau data masa lalu) kemudian mengkategorikan data tersebut kedalam golongan-golongan yang dikehendakinya.
2. Menghitung akar dari pohon. Atribut-atribut yang dipilih oleh analis disebut sebagai akar melalui proses perhitungan *gain* yang didapat karena adanya perubahan *entropy* pada setiap atribut tersebut. Apabila *gain* menunjukkan tertinggi, maka atribut tersebut akan dijadikan akar pokok atau paling atas. Nilai *gain* berasal dari nilai *entropy*. Berikut dirumuskan proses perhitungan nilai *entropy* pada sebuah atribut:

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

**Rumus 2.1** Perhitungan *Gain*



Rumus diatas, diartikan sebagai persamaan sistematis untuk menemukan nilai *gain* yang berasal dari *entropy*, dengan keterangan sebagai berikut:

- S : himpunan kasus
- A : atribut
- N : jumlah partisi atribut A
- |Si| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

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

### **Rumus 2.2** Perhitungan *Entropy*

Rumus diatas, diartikan sebagai persamaan sistematis untuk menemukan keandalan suatu hasil entropy yang telah dilakukan pada Rumus 2.1, dengan keterangan sebagai berikut:

- S: himpunan kasus
  - A : fitur
  - N : jumlah partisi S
  - Pi : proporsi dari Si terdapat S
3. Ulangi langkah ke-1 dan langkah ke-2 pada setiap atribut yang telah dipilih sampai seluruh *record* ikut berpartisipasi.
  4. Ada hal-hal yang harus diperhatikan yakni kapan partisi pohon keputusan harus dihentikan, yakni ketika:
    - a. Semua *record* dalam simpul N mendapat kelas yang sama.

- b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
- c. Tidak ada *record* di dalam cabang yang kosong.

Menurut pendapat dari Marwana (2013) bahwa terdapat dua hal pokok atau mendasar dari proses menemukan hasil *entropy*, antara lain.

1. Apabila pada kolom “benar” atau “tidak” menunjukkan nilai 0 (nol) dengan demikian nilai dari *entropy*-nya berbanding lurus atau akan menunjukkan nilai 0 (nol).
2. Apabila pada kolom “benar” dan “tidak” menunjukkan nilai yang sama dengan demikian nilai dari *entropy*-nya sudah pasti menunjukkan nilai 1 (satu).

Hasil nilai *gain* dan *entropy* yang dilakukan pengujiannya bisa menggunakan alat bantu yakni *software Waikato Environment for Knowledge Analysis (WEKA)*. Seperti yang telah dilakukan oleh (Rifqo & Arzi, 2016), menemukan bahwa algoritma C4.5 mendukung pembuatan pohon keputusan yang dapat dilakukan dengan cara-cara dibawah ini:

1. Pilih atribut sebagai akar, merupakan atribut yang tidak memiliki cabang masukan dan berpengaruh paling besar pada suatu kelas tertentu.
2. Buat cabang untuk tiap-tiap nilai, adalah membuat struktur cabang dari *decision tree* yang sudah berisi nilai-nilai.
3. Bagi kasus dalam cabang, adalah melakukan pembagian atau pengelompokkan dari permasalahan yang terdapat pada struktur cabang dari *decision tree*.

4. Ulangi proses untuk setiap cabang sampai kasus pada cabang memiliki kelas yang sama, adalah melakukan proses pengulangan yang ada pada tahapan pertama hingga ketiga sampai seluruh cabang pada struktur *decision tree* memiliki kelas dan kategori yang sama.

### **2.1.2 Teori Khusus**

Dijabarkan teori khusus yang diambil oleh peneliti, dimana penjelasannya dapat diuraikan seperti dibawah ini:

#### **2.1.2.1 Profit**

Menurut (Elisa, 2018), profit atau nama lainnya ialah laba, didefinisikan oleh Martono dan Harijito sebagai selisih antara pendapatan yang menjadi penghasilan yang diterima oleh perusahaan dari satu pihak maupun beberapa pihak dengan biaya yang menjadi beban yang harus dikeluarkan guna memenuhi kewajibannya kepada pihak lain. Profit atau laba adalah peningkatan nilai modal dari peristiwa-peristiwa masa lalu operasional perusahaan yang terjadi karena adanya transaksi-transaksi keluar masuknya uang yang memberikan pengaruh pada perusahaan dalam jangka waktu tertentu. Lain halnya dengan profit bersih atau laba bersih, didefinisikan sebagai selisih lebih atau bernilai positif dari pendapatan/penghasilan dengan beban/biaya kemudian dikurangi lagi dengan bunga dan pajak.

Berdasarkan pendapat (Elisa, 2018), Sistem yang dibangun akan dipergunakan untuk Prediksi Pencapaian Profit Pada Perusahaan. Data yang semula mentah diolah menjadi berbentuk angka maupun non-angka kemudian

diklasifikasikan menjadi beberapa kelas sehingga mempermudah penganalisisan data pada tahap berikutnya. Kemudian data tersebut dibentuk berdasarkan desain pohon keputusan sebagai output. Beberapa hal untuk mengambil keputusan dalam prediksi pencapaian profit pada perusahaan yakni seperti dibawah ini:

A. Biaya Langsung Personil

1. Biaya Tenaga Ahli
2. Biaya Tenaga Teknik
3. Biaya Tenaga Pendukung

B. Biaya Langsung Non Personil

1. Biaya Operasional Kantor
2. Biaya Peralatan
3. Biaya Pelaporan & Dokumen
4. Biaya Mobilisasi

Variabel yang ditentukan untuk diambil keputuannya ialah BAIK dan TIDAK BAIK.

#### **2.1.2.2 *Waikato Environment for Knowledge Analysis (WEKA)***

WEKA adalah ialah sebuah aplikasi dari *machine learning* praktis. “WEKA” berasal dari singkatan *Waikato Environment for Knowledge Analysis*, berasal dari sebuah perguruan tinggi, yakni Universitas Waikato, New Zealand digunakan pada penelitian, edukasi serta aplikasi-aplikasi lainnya. WEKA memiliki kemampuan untuk mengatasi berbagai permasalahan *data mining* di

dunia nyata, terutama pengklasifikasian dengan dasar pendekatan *machine learning*. Aplikasi ini ditulis dalam hirarki *class Java* dengan metode berorientasi objek dan dapat berjalan hampir di semua platform.

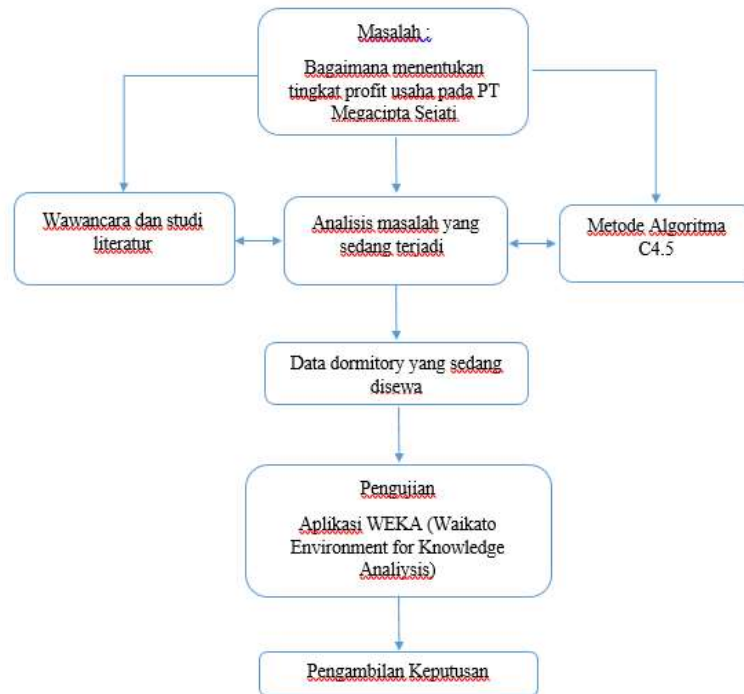


**Gambar 2.6** Tampilan Utama WEKA

**Sumber:** (Witten, 2011)

## 2.2 Kerangka Berpikir

Kerangka berpikir merupakan penyelesaian masalah dari penelitian yang ada berdasarkan langkah-langkah yang telah disusun, kerangka berpikir yang telah disusun seperti dibawah ini:



**Gambar 2.7** Kerangka Berpikir

**Sumber:** (Hasil Penelitian, 2020)

### 2.3 Hipotesis Penelitian

hipotesis merupakan pernyataan yang berisi gambaran atau prediksi adanya hubungan antara dua atau lebih variabel, yang pembuktian kebenarannya masih harus melalui beberapa pengujian. Berdasarkan uraian di atas, dirumuskan hipotesis sebagai berikut :

1. Teknik *datamining* yaitu teknik klasifikasi dengan pohon keputusan menggunakan algoritma C4.5 dapat menganalisis data profit usaha.
2. Hasil penelitian dapat membantu menentukan tingkat profit usaha pada PT Megacipta Sejati.

## 2.4 Penelitian Terdahulu

Berikut diuraikan beberapa penelitian yang telah dilakukan oleh peneliti terdahulu yang dijadikan sebagai acuan dari penelitian ini. Penelitian terdahulu bertujuan agar penelitian yang dilakukan memperoleh hasil yang optimal dan valid. Berikut adalah beberapa penelitian yang dijadikan acuan.

1. Anu And Chauhan Harvinder (2013) penelitian yang berjudul “Implementation Of Decision Tree Algoritma C4.5”. Penelitian ini menunjukkan algoritma ini lebih cocok untuk data sets yang mengandung tidak terlalu banyak data karena hasil yang lebih akurat dan pengimplementasian pohon keputusan dengan metode algoritma C4.5 sangat mudah dan cepat.
2. Jambur (2016) penelitian yang “Menganalisa jumlah pelanggan aktif dengan menerapkan teknik *datamining* metode algoritma C4.5”. Hasil yang didapat mampu membantu dalam memfilter pelanggan yang aktif dan tidak aktif pada perusahaan yang menunjukkan bahwa pelanggan dengan jumlah bayar tinggi, pembelian pertahun sedang, diskon rendah maka merupakan pelanggan aktif dan sebaliknya maka merupakan pelanggan pasif.
3. Pujiono et al., (2013) penelitian tentang “Penggunaan *WEKA* untuk menganalisis kepuasan publik dalam mewujudkan *good governance* di kota Yogyakarta”. Penelitian ini menunjukkan bahwa pelaksanaan *good governance* dapat disimpulkan akan berjalan dengan baik karena pelayanan yang diberikan Dinas Perizinan selama ini cukup baik.

4. Harman (2018) penelitian tentang “Memprediksi indikator *website* yang baik dengan menerapkan teknik *datamining* metode algoritma C4.5”. Penelitian ini menunjukkan bahwa indikator navigasi dan graphic design adalah indikator yang paling mempengaruhi indikator *website* yang baik.
5. Mardin (2017) penelitian tentang “klasifikasi menggunakan algoritma C4.5”. Penelitian ini menunjukkan bahwa algoritma C4.5 dapat menghasilkan pohon keputusan dengan metode klasifikasi yang dapat memberikan informasi yang dibutuhkan.
6. Chen, Zhang, Li, & Shahabi, (2018) penelitian yang berjudul “*Perfomance evaluation of the GIS-based datamining techniques of best-first decision tree, random forest, and naive bayes tree for landslide susceptibility modeling*”. Penelitian ini menunjukkan bahwa teknik data mining yang paling akurat dalam memprediksi merupakan teknik *random forest model*.
7. Elisa (2018) penelitian tentang “Memprediksi profit pada perusahaan dengan menggunakan teknik *datamining* metode algoritma C4.5”. Hasil penelitian mampu menjadi sebuah referensi bagi perusahaan yang ingin meningkatkan profitnya dengan cara *saving cost*. Hasil penelitian ini merupakan biaya yang paling banyak mempengaruhi profit perusahaan adalah biaya mobilisasi dan biaya tenaga kerja.
8. Azwanti (2018) penelitian tentang “Memprediksi mahasiswa yang mengulang mata kuliah menggunakan teknik *datamining* metode algoritma C4.5. Penelitian ini dilakukan di AMIK Labuhan Batu



menyimpulkan bahwa algoritma C4.5 merupakan algoritma yang sangat membantu dalam mengklasifikasi data karena dapat memperoleh dengan jelas karakteristik data yang diklasifikasi baik dalam bentuk pohon keputusan maupun *if-then*.