

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Teori Umum**

Penulis membahas beberapa teori umum berkaitan dengan pelaksanaan penelitian terdiri dari pengertian dan istilah secara umum.

##### **2.1.1. KDD (Knowledge Discovery In Database)**

Metode penciptaan pengetahuan baru digunakan untuk mendapatkan informasi disimpan dalam basis data adalah KDD, merupakan singkatan dari Knowledge Discovery In Database. Jika informasi baru ditemukan, itu dapat digunakan sebagai dasar pengambilan keputusan. Langkah-langkah KDD dijelaskan di bawah ini:

1. *Data Selection*

Proses pertama harus dilakukan adalah melakukan kegiatan pengumpulan data dengan melalui langkah-langkah pemilihan data agar nantinya kumpulan data tersebut dapat diolah.

2. *Pre-processing* atau *Cleaning*

Setelah pemilihan data selesai maka dapat dilanjutkan dengan pembersihan data dari adanya duplikasi data yaitu pengecekan data diperlukan.

3. *Transformation*

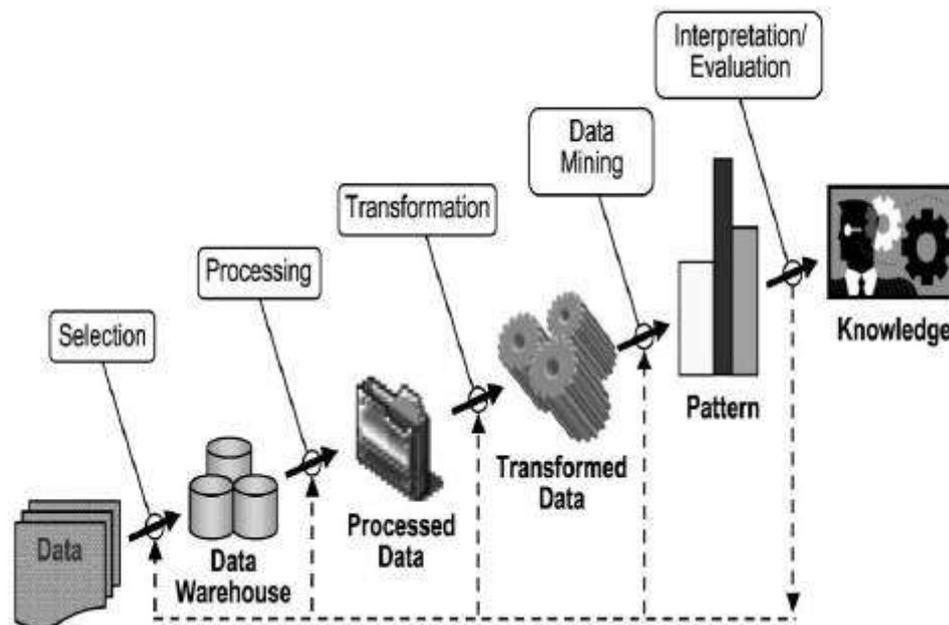
Langkah selanjutnya setelah pembersihan data adalah transformasi data, di mana siklus transformasi data dilakukan dan dibandingkan dengan database disimpan.

#### 4. *Data Mining*

Tahap selanjutnya adalah data mining, dimana informasi dicari hingga terbentuk sistem menjadi tujuan penelitian untuk mengambil keputusan.

#### 5. *Interpretation*

Langkah terakhir harus dilakukan adalah interpretasi atau pola dihasilkan dari proses sebelumnya, dapat dipahami dengan melihat pola dihasilkan.

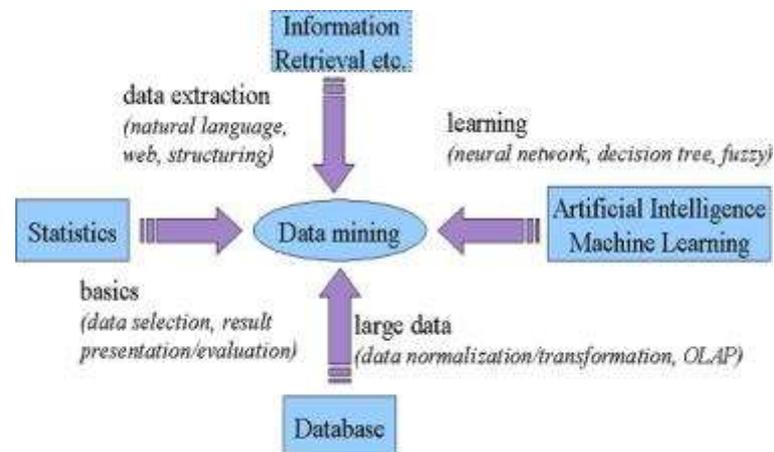


**Gambar 2. 1** Proses Dalam KDD atau *Knowledge Discovery in Database*

**Sumber:** (Mardi, 2016)

### 2.1.2. Data Mining

Di setiap jalur penambangan data, suatu teknik dapat digunakan untuk memproses data sehingga dapat membentuk skema itu sendiri, menggunakan skema terbentuk untuk mengidentifikasi skema lain tidak disimpan di Pusat Database. mencakup informasi dan pengumpulan data serupa. Penambangan data bekerja sangat baik sebagai penggunaan implementasi fase peramalan. Asal muasal bidang data mining adalah pengolahan data sudah cukup lama diproses, seperti B. Mining agar data ada dapat diolah untuk menghasilkan informasi berguna dan bernilai. Sebuah cabang ilmu data mining juga dapat disebut sebagai mesin pembelajaran dapat digunakan untuk fungsi membentuk beberapa analisis, dimana pembelajaran sederhana dapat dijelaskan dan diterima oleh pikiran manusia dengan cara menghasilkan pengetahuan baru. (Yuli Mardi, 2019).



**Gambar 2. 2** Bidang Ilmu Data mining

**Sumber:** (Yuli Mardi, 2019)

### 2.1.3. Pengelompokan *Data Mining*

Ketika sebuah fase didasarkan pada tugas dan proses dilakukan dalam penambangan data, fase itu dapat dibagi menjadi beberapa fase dan bagian.

#### 1. *Description* atau Deskripsi

Pada titik inilah analis mulai mencoba mengklasifikasikan pola dalam data dan sumber data. Hal ini disebabkan oleh fakta bahwa penanggung jawab penghitungan suara dalam kampanye pemilu tidak mengetahui atau memiliki informasi jelas tentang keterampilan, kemampuan, dan kemampuan kandidat untuk berpartisipasi dalam pemilu. Oleh karena itu, penjelasan masalah merupakan upaya untuk memberikan penjelasan tentang potensi kemampuan seorang calon untuk mendaftar mengikuti pemilu dengan rencana.

#### 2. *Estimation* atau Estimasi

Juga, kegiatan dilakukan oleh evaluator dalam langkah atau proses ini serupa dengan proses atau langkah klasifikasi, tetapi variabel objektif terlibat dalam mengevaluasi proses atau langkah melibatkan data digital mungkin berbeda. Sampel dihasilkan berasal dari dataset berisi nilai atau angka variabel digunakan sebagai dokumen untuk mendapatkan dari proses peramalan. Selanjutnya tentukan estimasi diturunkan dari nilai prediksi dari nilai-nilai terdapat pada variabel turunan. Hal ini dapat dijelaskan dengan nilai (S.1) terkandung dalam rata-rata IPK kumulatif atau rata-rata IPK kumulatif mahasiswa sedang menempuh pendidikan tinggi. Kelayakan

ditentukan berdasarkan nilai IPK atau Indeks Prestasi Mahasiswa. Proses pendidikan untuk setiap semester.

### 3. *Prediction* atau Prediksi

Proses atau langkah peramalan mirip dengan proses klasifikasi dan evaluasi, tetapi tidak melibatkan peramalan apakah nilai saat ini akan menjadi nilai masa depan. Prediksi disebutkan sangat relevan untuk penelitiandan bisnis tepatnya,

- a. Saat mereka membuat ramalan untuk mengetahui kemungkinan harga beras untuk beberapa bulan mendatang.
- b. Pada saat sama, ada perkiraan kemungkinan tingkat pengangguran dalam beberapa tahun ke depan.
- c. Perkiraan dalam persentase menggambarkan kemungkinan kecelakaan jika pengemudi melebihi batas kecepatan di tahun-tahun mendatang.

### 4. *ClassificatIon* atau Klasifikasi

Suatu proses atau fungsi akan diranking memiliki tujuan berbeda dengan variabel berdasarkan kategori ada. Hal ini dapat dijelaskan dengan fakta bahwa ketika mengklasifikasikan pendapatan karyawan perusahaan, pendapatan karyawan dibagi menjadi tiga kategori. Dengan kata lain, pekerja kelompok pendapatan adalah pekerja dengan kelompok pendapatan sangat tinggi, menengah atau biasa. Pekerja sangat rendah Jika klasifikasi terakhir diartikan dalam hal penelitian dan bisnis:

- a. Ada transaksi kredit curang atau curang dalam proses penilaian.
- b. Ketika pelanggan mengantisipasi rencana kredit, mereka melakukannya terlepas dari apakah itu benar atau tidak.
- c. Saluran Dokter mendiagnosis penyakit dan penyakit dialami dan didapat oleh pasien berdasarkan gejala muncul.

#### 5. *Clustering* atau Pengklusteran

Dokter mendiagnosis penyakit dan kondisi dialami dan didapat oleh pasien berdasarkan gejala mereka tunjukkan. Cluster bersangkutan memiliki hasil input sama atau mirip dengan kelompok sampel, dan kelompok lain tidak sama atau mirip dengan hasil input ada. Satu atau lebih proses terlibat dalam pengelompokan berbeda dari proses klasifikasi atau satu atau lebih langkah karena pengelompokan tidak memiliki variabel target. Pada saat menghitung Algoritma digunakan pada proses pengelompokan, penekanannya adalah pada pengelompokan data ada berdasarkan kesamaan pola bahan, sehingga pekerjaan dilakukan pada tahapan dan proses pengelompokan adalah perkiraan. , Proses prediksi dan klasifikasi tidak dilakukan alokasi. Nilai maksimum muncul dari atau menurut kesamaan atau ketidaksamaan data dalam kelompok. Jika kesamaan atau kemiripan dengan data ada sangat sedikit, maka terjadi adalah sebaliknya. Nilai dihasilkan Ini adalah nilai minimum. Hal ini dapat dijelaskan sebagai berikut.

2.1.4. Saat ini, kelompok konsumen menjadi sasaran kegiatan pemasaran produk perusahaan.

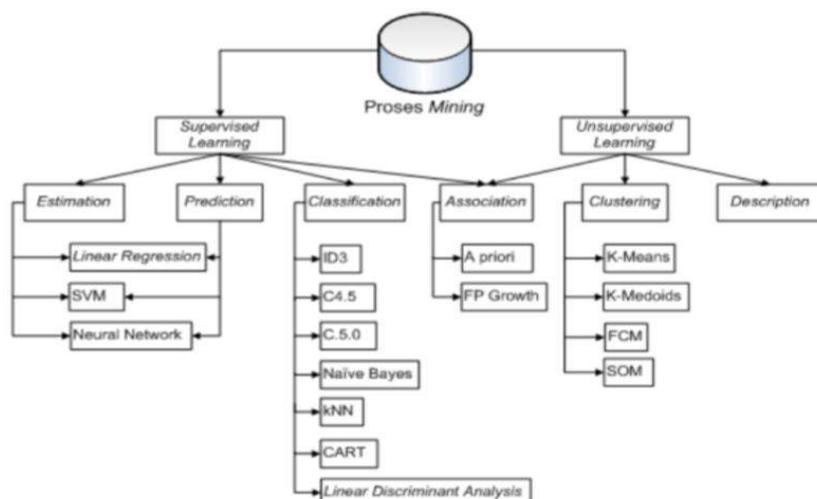
2.1.5. Saat melakukan pemeriksaan, kami membedakan antara kami anggap mencurigakan dan kami anggap pasti.

#### 6. *Association* atau Asosiasi

Asosiasi mengimplementasikan proses untuk memulihkan integritasi data atau sebagai pendukungnya. Hal ini dapat dijelaskan dengan bentuk tindakan dilakukan,

- Selama survei atau analisis basis pelanggan perusahaan, mereka memberikan umpan balik positif tentang layanan diberikan.
- Cari tahu tentang semua jenis produk di pasaran belum pernah dibeli atau dikonsumsi.

Interpretasi visual melibatkan pengelompokan proses tertentu atau hasil langkah demi langkah ketika peneliti melakukan penelitian menggunakan teknik data mining seperti:

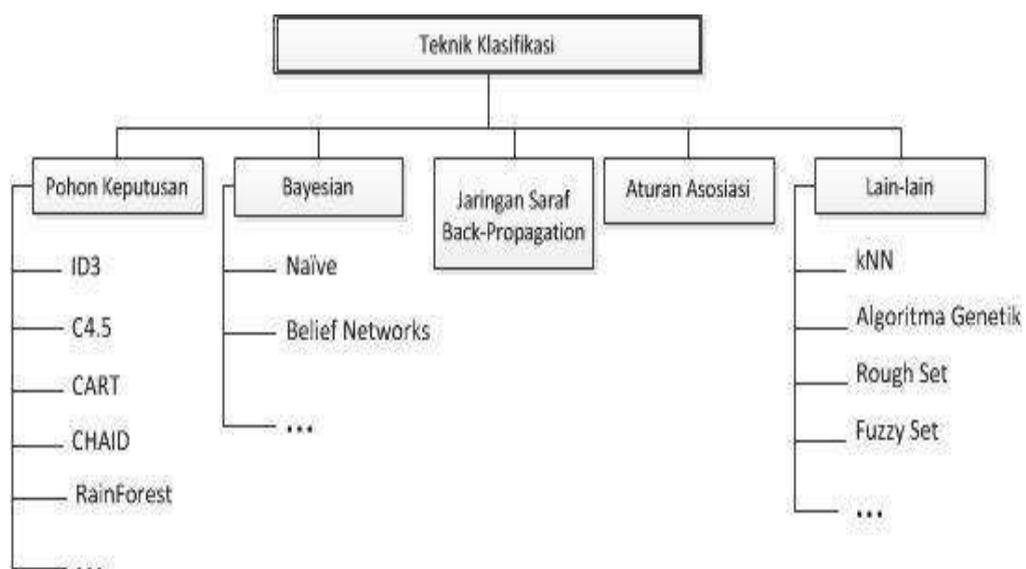


**Gambar 2. 3** Pengelompokan Teknik Data mining

**Sumber:** (Hermawan, Sukma, & Halfis, 2019)

#### 2.1.4. Klasifikasi

Klasifikasi juga dapat digambarkan sebagai kegiatan melakukan pengumpulan sebagai diperlukan dari beberapa objek akan diuji kemudian menjelaskan beberapa objek tersebut untuk mendapatkan informasi diperlukan. Proses klasifikasi dilakukan untuk menentukan rencana pelatihan memetakan beberapa atribut ke dalam kelompok telah ditentukan, sehingga atribut tersebut dapat dijelaskan dari kelompok sebelumnya tidak dapat dikenali. Seperti dikemukakan oleh (Azwanti, 2018), klasifikasi sebagai dapat dilihat sebagai kegiatan mengumpulkan alat-alat serupa dalam sekelompok objek dari sebagai base dan mengklasifikasikannya ke dalam kelompok tertentu sesuai dengan harapan analisis. Tujuan klasifikasi adalah untuk mendefinisikan model set pelatihan mengurutkan atribut ke dalam kelas sesuai, kemudian mengklasifikasikan atribut tersebut ke dalam kelas tidak ada sebelumnya. Ada beberapa teknik klasifikasi dapat digunakan seperti terlihat pada gambar di bawah ini.

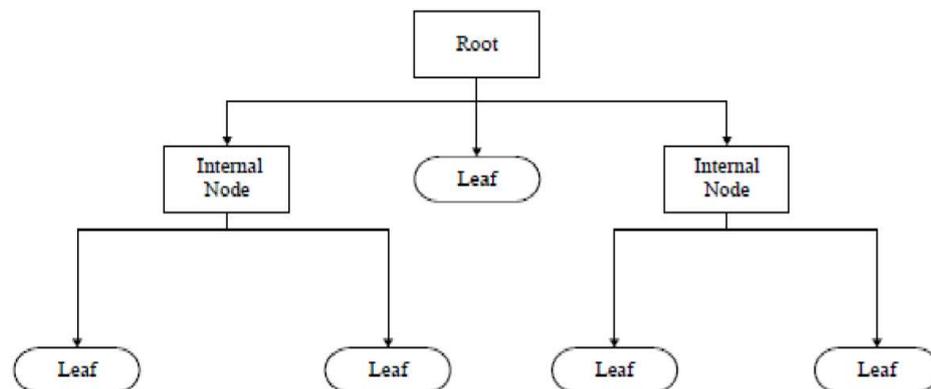


**Gambar 2. 4** Pengelompokan Teknik Klasifikasi

**Sumber:** (Della, 2020)

### 2.1.5. Decision Tree

Pohon keputusan memberikan solusi dan memecahkan berbagai masalah melalui keputusan absolut dan dapat digunakan dimana sering membantu dalam pengambilan keputusan di bidang Ilmu data mining. Pohon keputusan sangat efektif bila digunakan sebagai alat untuk mengubah informasi ada menjadi model menciptakan prinsip-prinsip mudah dipahami dan dapat dicerna secara logis. Dalam prakteknya, pohon keputusan menerapkan kegiatan diawali dengan pengumpulan informasi secara jelas mencakup keadaan dan latar belakang setiap individu dan dapat mengambil keputusan untuk masa depan. (Azwanti, 2018).



Gambar 2. 5 Model Decision Tree

**Sumber:** (Takalapeta, 2018)

### 2.1.6. Algoritma C4.5

Dengan menggunakan Algoritma, lebih mudah untuk membuat sistem pendukung keputusan berupa pohon akan membantu dalam pengambilan keputusan. Algoritma c4.5 juga dapat dikatakan sebagai perubahan berhasil diterapkan pada Algoritma ID3 hingga tercapainya Algoritma c4.5, sehingga

Algoritma c4.5 juga dapat dikatakan sebagai Algoritma turunan dari Algoritma ID3. Algoritma C4.5 memecahkan masalah ada dan memberikan solusi untuk kehilangan sebagai dan solusi distribusi sebagai. Alur pembuatan himpunan terstruktur sering dilakukan dengan menggunakan Algoritma c4.5 untuk menghasilkan pohon keputusan (Fikri & Verina, 2020).

$$Gain(S, A) = Entropi(s) - \sum_i^n \frac{|S_i|}{|S|} \times Entropi(S_i)$$

### Rumus 2. 1 Perhitungan Gain

Keterangan dari rumus 2.1 diatas dapat dijabarkan sebagai berikut::

$S$  berupa himpunan kasus

$A$  berupa attribute digunakan.

$n$  berupa jumlah partisi dimiliki oleh attribute  $A$

$|S_i|$  berupa jumlah dimiliki oleh partisi ke- $i$

$|S|$  berupa jumlah dimiliki oleh partisi  $S$

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

### Rumus 2. 2 Perhitungan Entropy

Keterangan dari rumus 2.2 diatas dapat dijabarkan sebagai berikut::

$S$  berupa himpunan kasus

$A$  berupa attribute digunakan

$n$  berupa jumlah partisi dimiliki oleh attribute  $S$

$p_i$  berupa proporsi dari  $S_i$  kepada  $S$ .

## **2.2. Teori Khusus**

Penulis membahas beberapa teori penelitian khusus terdiri dari definisi dan istilah tertentu.

### **2.2.1. Penjualan**

Penjualan adalah pembuatan pertukaran real estat untuk uang, di mana perusahaan memasok barang sehingga perusahaan dapat menerima uang dari orang menginginkan properti tersebut. Perusahaan memperdagangkan barang dan jasa juga dapat disebut sebagai badan bergerak dalam bisnis jual beli, juga dapat dipahami sebagai aset diperoleh dapat diproses oleh badan tersebut di atas untuk kemudian didistribusikan atau didistribusikan (Sandrawira Anggraini, Sarjon Defit, 2018).

### **2.2.2. Profit**

Laba, sering disebut laba, adalah jumlah pemotongan, disebut juga pendapatan, diperoleh badan dengan cara menerimanya dari pihak luar menjadi tanggungan dan harus ditanggung untuk memenuhi kewajibannya kepada pihak luar. Peristiwa sebelumnya dalam operasi perusahaan, terjadi pada tanggal tidak ditentukan karena aktivitas perkreditan keuangan terkait keagenan, terjadi sebagai peningkatan modal, peristiwa ini disebut juga laba atau keuntungan. berbeda dengan keuntungan atau penghasilan bersih diterima, tetapi harus dikurangkan dari biaya-biaya seperti bunga dan pajak dibayarkan (Elisa, 2018).

### 2.2.2. Walkato Environment for Knowledge Analysis (WEKA)

Kumpulan tetap bisa disebut repositori Java dari berbagai tingkatan digunakan untuk meningkatkan kumpulan sebagai pada mesin dapat digunakan sebagai alat penelitian untuk algoritme seperti penambangan sebagai juga bisa disebut WEKA.

Gudang bertingkat, sebagian besar terkomputerisasi dengan aplikasi online, sangat memudahkan penggunaan mesin belajar untuk informasi dari operasi komputer(Azwanti, 2018).



**Gambar 2. 6** Tampilan Utama *WEKA*

### 2.3. Penelitian Terdahulu

Peneliti juga menemukan beberapa hasil dari penelitian sebelumnya terkait dengan analisis menggunakan Algoritma C4.5 yaitu:

1. Dari penelitian “Prediksi Waktu Kuliah Dengan Algoritma Classifier” dilakukan oleh (Rika Harman & Evan Rosiska, 2020) Didapatkan informasi bahwa dengan menggunakan Algoritma tipe C4.5 dapat membantu mahasiswa

menemukan akar permasalahan dan mengidentifikasi penyebab utama mengapa mahasiswa tidak dapat menyelesaikan mata kuliahnya tepat waktu.

2. Dari penelitian "*Taiga: Performance Optimization of the C4.5 Decision Tree Construction Algorithm*" dilakukan oleh ( & Chen, 2016) Diperoleh informasi bahwa penggunaan Algoritma tipe C4.5 pada pohon keputusan dapat dimaksimalkan dengan menyelaraskan sebagai dengan kategori atau kriteria dalam data mining sehingga waktu diperlukan untuk membuat keputusan akhir dapat dikurangi.
3. Dari penelitian "*Penerapan Algoritma C4.5 Untuk Menentukan Reward Karyawan Pada PT Indoland Batam*" dari (Harman, 2021) Ditetapkan bahwa penelitian ini dapat mempercepat proses perhitungan atau menentukan penghargaan tahunan lebih efisien.
4. Dari penelitian "*Performance improvement of sebgaimining ini WEKA through GPU acceleration*" dilakukan oleh (Augusto & Char, 2014) melaporkan bahwa proses memakan waktu dapat beradaptasi dengan GPU. *show that set of operations that are time consuming can easily adapted to GPUs.* Ketika memparalelkan metode perkalian matriks, dapat menyesuaikan penggunaan CPU

dan GPU sesuai dengan ukuran matriks, metode ini bahkan mempercepat prosesnya 49%.

5. Dari penelitian “Penerapan Algoritma C.45 Untuk Kelayakan Penerima Uang Dan Beras” dilakukan oleh (Harman, 2018) Dilaporkan bahwa penggunaan algoritme tipe C4.5 diharapkan dapat meningkatkan praktik sebelumnya dianggap tidak adil atau, menurut penelitian ini, faktor kedekatan akan lebih penting bagi penerima uang tunai dan bantuan perjalanan di masa mensebagaing.
6. Dari penelitian “Analisis Dengan Metode Klasifikasi Menggunakan *Decission Tree* Untuk Prediksi Penentuan Resiko Kredit Pada Bank Bukopin Batam” dilakukan oleh (Intan Utna Sari, 2015) Informasi diterima bahwa penggunaan Algoritma tipe C4.5 akan membantu bank mempercepat dan mempermudah memprediksi sebagai dan menarik kesimpulan dari sebagai ada. Salah satu cara untuk memprediksi sebagai adalah dengan menggunakan software Dtreg dan Decision Tree. Dtreg mampu membuat pohon keputusan salah satunya adalah hasil keputusan risiko jumlah KPR berdasarkan jumlah nasabah sehingga diharapkan dengan teknik decision gap ini akan membantu bank menjadi lebih cepat dan mudah diprediksi. dan menarik kesimpulan dari sebagai ada.

7. Dari penelitian “Penerapan Algoritma C4.5 Untuk Memprediksi Penjualan Sayur Pada PT Bumi Bhakti Barelang” dari (Harman, 2022) Penelitian ini terbukti dapat membantu dalam memfasilitasi pengiriman sayur sop di PT Bhakti Bumi Barelang dengan merencanakan pengiriman sayur sop dan menginformasikan perusahaan tentang produk paling banyak dibeli oleh konsumen.
8. Survei “Prediksi Profit Perusahaan Menggunakan Algoritma C4.5” dilakukan oleh (Elisa, 2018) jelas menunjukkan bahwa penggunaan Algoritma C4.5 efektif. Untuk memprediksi profitabilitas perusahaan, dimungkinkan untuk membangun sistem pendukung keputusan dengan standar diberikan.
9. Dari penelitian “Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT. Capella Dinamik Nusantara Cabang Muka Kuning dilakukan oleh (Azwanti, 2018) Ditemukan bahwa Algoritma tipe C4.5 berguna dalam kegiatan atau proses melakukan klasifikasi dan dapat juga digunakan untuk memprediksi penjualan sepeda motor. Selain itu juga memudahkan peneliti dalam tahap pengujian atau proses dengan aplikasi WEKA.
10. Setelah (Rika Nofitri, 2018) melakukan penelitian “Analisis Kinerja Algoritma C4.5 dalam Prediksi Kinerja Laba”, Ditemukan bahwa Algoritma tipe C4.5 berguna dalam

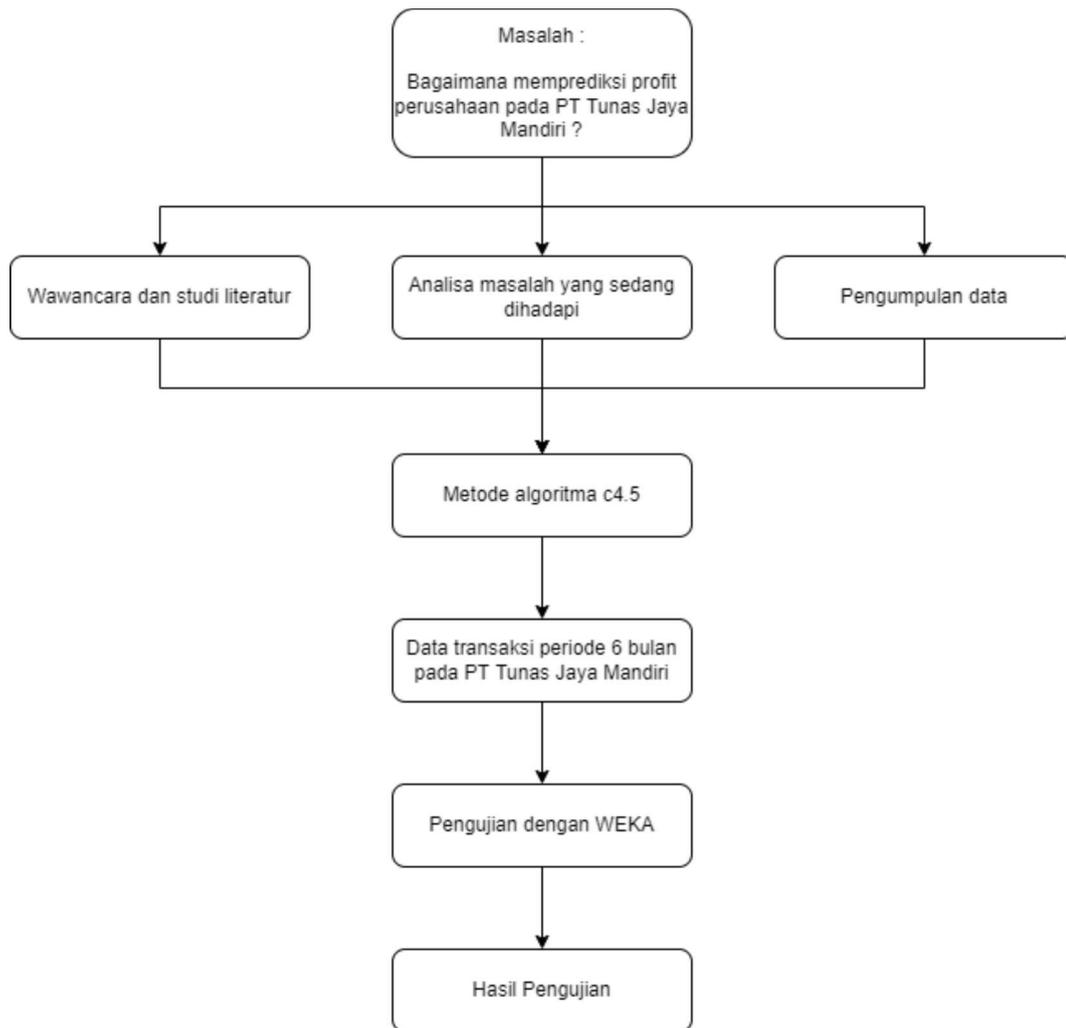
kegiatan atau proses melakukan klasifikasi dan dapat juga digunakan untuk memprediksi penjualan sepeda motor. Selain itu juga memudahkan peneliti dalam tahap pengujian atau proses dengan aplikasi WEKA.

11. Setelah (Harman, 2020) melakukan survey “Pada penerapan Algoritma C4.5 untuk menentukan tingkat keuntungan perusahaan di PT Megacipta Sejati, didapatkan hasil bahwa penggunaan Algoritma C4.5 dapat dibangun. Suatu sistem pendukung keputusan untuk mengetahui tingkat kelangsungan usaha perusahaan sesuai dengan kriteria telah ditetapkan oleh perusahaan.
12. Penelitian oleh (Tukino, 2019) tentang "Penerapan Algoritma C4.5 untuk memprediksi keuntungan di PT. SMOE Indonesia" telah menunjukkan bahwa penggunaan Algoritma C4.5 dapat membuat bantuan keputusan. Sistem peramalan keuntungan perusahaan didasarkan pada kriteria berikut Ditentukan oleh perusahaan.
13. Dari penelitian “Analisis Algoritma C.45 Untuk Menentukan Faktor Pembelian Sepeda Bekas Pada Toko Sepeda Batam” dari (Harman, 2021) Teramati bahwa penerapan Algoritma C4.5 akan membantu perusahaan untuk mempelajari pilihan pembeli saat membeli sepeda bekas.

14. Penelitian tentang memprediksi mahasiswa mengulang mata kuliah menggunakan teknik sebagaimining metode Algoritma C4.5 ini dilakukan di AMIK Labuhan Batu dan diteliti oleh (Azwanti, 2018) menyimpulkan bahwa Algoritma C4.5 merupakan Algoritma sangat membantu dalam mengklasifikasi sebagai karena dapat memperoleh dengan jelas karakteristik sebagai diklasifikasi baik dalam bentuk pohon keputusan maupun if-then
15. Dari penelitian“Penerapan Data mining Untuk Prediksi Penjualan Readymix Menggunakan Metode Algoritma C4.5 Pada PT Remicon Widyaprima” dari (Harman, 2020) didapati penelitian ini membahas mengenai prediksi penjualan PT Remicon Widyaprima agar mengetahui class paling banyak diminati oleh konsumen ada di kota batam. Dimana setiap class pada penjualan readymix memiliki keunggulan masing-masing.

#### **2.4. Kerangka Pemikiran**

Berikut adalah kerangka pemikiran dalam penelitian ini:



**Gambar 2. 7** Kerangka Berpikir