

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori Dasar**

Peneliti berupaya menjelaskan mengenai teori yang menjadi landasan dari kegiatan penelitian ini dilaksanakan, dengan mendefinisikan ide penelitian dan melakukan pengkajian terhadap teori yang akan digunakan.

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

Peneliti berupaya menyatakan bahwa penelitian ini dilakukan dengan benar dan bersifat universal dan dapat dipergunakan sebagaimana mestinya dengan menjelaskan teori yang berhubungan dengan penelitian ini menggunakan teori yang bersifat umum.

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

Suatu cara yang digunakan untuk menghasilkan pengetahuan baru yang dimanfaatkan untuk mendapatkan data yang telah disimpan kedalam basis data adalah KDD yang merupakan singkatan dari Knowledge Discovery In Database. Setelah pengetahuan baru ditemukan, pengetahuan tersebut dapat dimanfaatkan sebagai dasar yang mendukung dalam mengambil keputusan. Pada berikut ini akan dijelaskan tahapan dalam KDD yaitu:

1. Data Selection

Proses awal yang akan dilaksanakan adalah melaksanakan kegiatan untuk mengumpulkan data dengan melewati tahapan seleksi data agar sekumpulan data tersebut dapat diproses pada tahap selanjutnya.

## 2. Pre-processing atau Cleaning

Setelah seleksi data dilakukan, dapat dilanjutkan dengan melakukan pembersihan data terhadap beberapa data yang terduplikasi yaitu dengan melakukan pemeriksaan terhadap data yang diperlukan.

## 3. Transformation

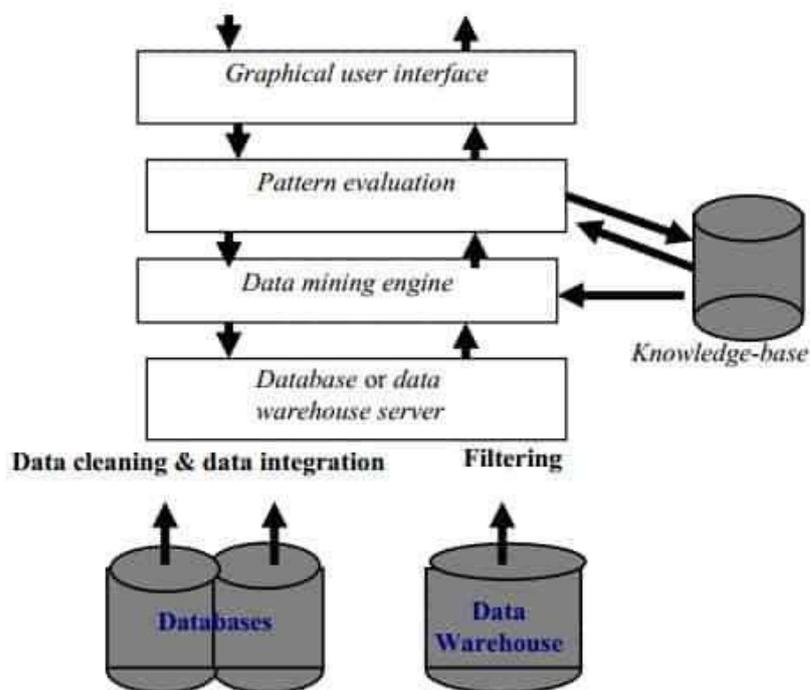
Tahapan selanjutnya yang akan dilakukan setelah melakukan pembersihan data adalah melakukan transformasi data dimana dilakukan siklus perubahan data dan disesuaikan dengan basis data yang telah disimpan.

## 4. Data Mining

Tahap berikutnya adalah tahap data mining dimana terjadi pencarian informasi hingga pembentukan system yang merupakan tujuan dari penelitian dilaksanakan hingga menghasilkan keputusan.

## 5. Interpretation

Tahap terakhir yang harus dilakukan adalah interpretasi atau pola yang dihasilkan dari proses sebelumnya dapat dipahami dengan menampilkan pola yang dihasilkan.



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

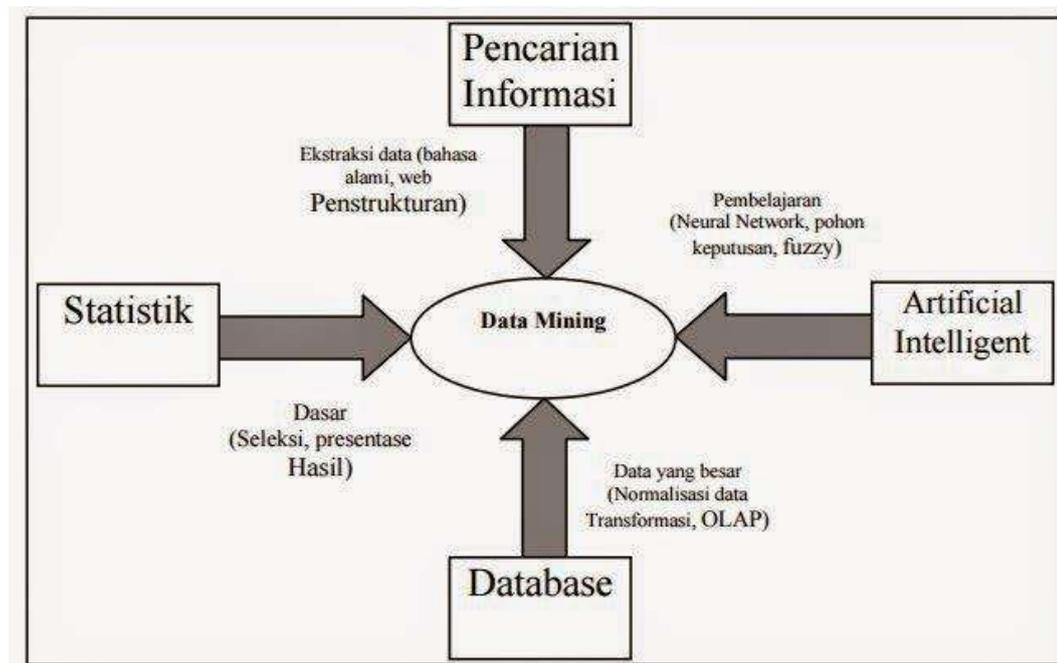
**Sumber:** (Hermawan et al., 2019).

### 2.1.1.2 Data Mining

Pada setiap alur dalam bidang ilmu data mining dapat digunakan suatu cara untuk mengolah informasi sampai dengan sendirinya dapat membentuk suatu skema, dimana skema yang terbentuk digunakan agar dapat mengidentifikasi skema lainnya yang belum terekam pusat database yang berisi informasi dan pendataan yang serupa. *Datamining* banyak difungsikan sebagai pemanfaatan terhadap terlaksananya langkah dalam memprediksi. Asal usul bidang ilmu data mining merupakan suatu pemrosesan informasi yang telah diproses cukup lama layaknya

terjadi kegiatan pertambangan untuk mengolah data yang ada untuk menghasilkan suatu informasi yang penting dan dapat dimanfaatkan. Bidang ilmu *datamining* yang dapat dikatakan juga sebagai mesin pembelajaran yang dapat digunakan untuk kegiatan membentuk sejumlah analisa dimana dapat dijelaskan pembelajaran yang sederhana dan dapat diterima oleh akal manusia hingga terciptanya ilmu pengetahuan yang baru (Yuli Mardi, 2019).

*Datamining* dapat dipahami sebagai langkah yang diambil untuk menemukan korelasi yang bermakna, statistik dan level tertentu dengan memilih di antara sekian data dan setelah itu mengumpulkannya pusat bank data berdasarkan kemajuan bidang ilmu statistika atau aritmatika. Bidang ilmu ini dapat dikatakan juga sebagai ilmu yang bekerja sama dengan bidang ilmu lainnya yaitu mesin pembelajaran, pengenalan pola, penyimpanan data, perwujudan dalam menyelesaikan permasalahan yang dihadapi pada saat pemrosesan data sehingga penerimaan data dapat dijadikan sebagai media untuk menghasilkan keputusan (Tusarwenda, 2018).



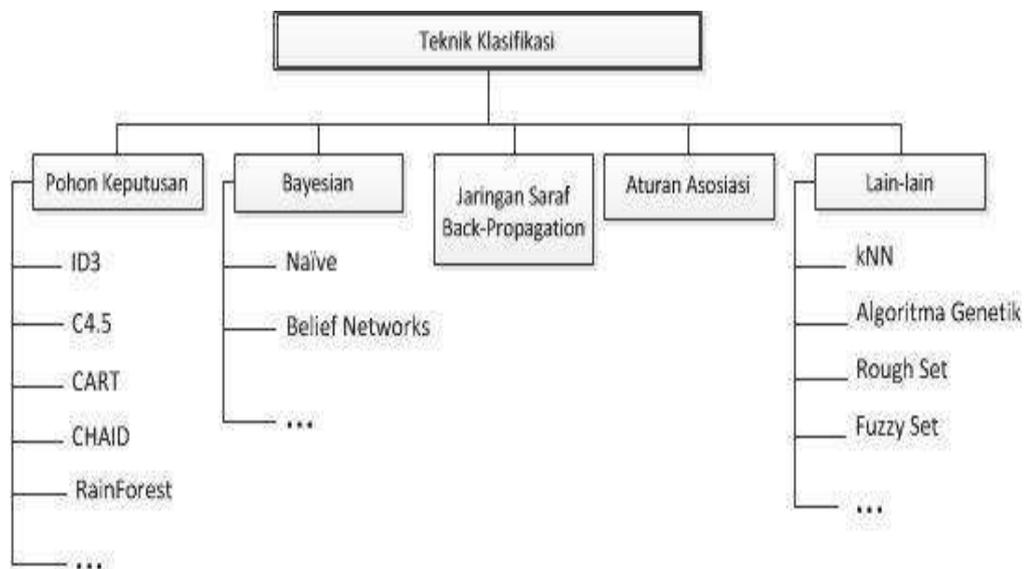
**Gambar 2.2** Bidang Ilmu Data Mining

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

### 2.1.1.3 Klasifikasi

Klasifikasi dapat disebut juga sebagai kegiatan dalam melaksanakan pengumpulan data yang diperlukan dari beberapa objek yang akan dijadikan sampel kemudian akan dilakukan pengklarifikasian terhadap beberapa objek tersebut hingga menghasilkan informasi yang dibutuhkan. Metode klasifikasi dilakukan agar dapat menentukan skema pelatihan yang menguraikan beberapa atribut kedalam kelompok yang telah ditentukan, kemudian atribut tersebut dapat dilakukan klarifikasi terhadap kelompok yang sebelumnya tidak dapat teridentifikasi. Seperti yang dikemukakan oleh (Azwanti, 2018), klasifikasi data

dapat dipandang sebagai kegiatan mengumpulkan alat-alat serupa ke dalam sekelompok objek yang diambil dari database dan mengklasifikasikannya ke dalam kelompok-kelompok tertentu sesuai dengan harapan analis. Klasifikasi bertujuan untuk menentukan model training set yang mengurutkan atribut-atribut ke dalam kelas-kelas yang sesuai, kemudian atribut-atribut tersebut diklasifikasikan ke dalam kelas-kelas yang sebelumnya tidak ada. ada beberapa teknik klasifikasi yang dapat digunakan seperti yang dapat dilihat pada gambar dibawah ini.



**Gambar 2.3** Pengelompokan Teknik Klasifikasi

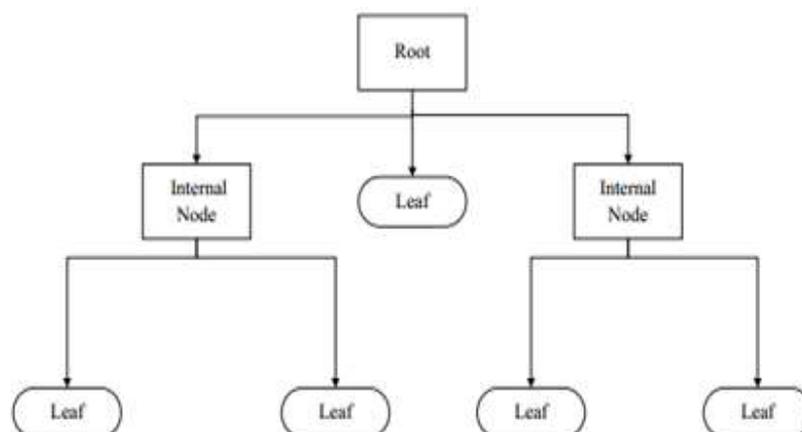
**Sumber:** (Della, 2020)

#### 2.1.1.4 Pohon Keputusan

Pohon keputusan memberikan solusi dan menyelesaikan berbagai permasalahan dengan keputusan yang mutlak dan dapat digunakan dimana hal tersebut menjadikan pohon keputusan sering digunakan dalam membantu

pengambilan keputusan sesuai dengan bidang ilmu data mining. Pohon keputusan cukup efektif bila dipergunakan sebagai media dalam mengkonversikan informasi yang ada menjadi suatu pola terciptanya prinsip yang mudah dimengerti dan dicerna dengan logika. Pada penerapannya, pohon keputusan mekaskanakan kegiatan yang berawal dari perkumpulan setiap bagian informasi yang mencakup keadaan dan latar belakang dengan jelas dan dapat mengambil suatu keputusan diwaktu mendatang (Azwanti, 2018).

Pohon keputusan (*decision tree model*) dapat diartikan juga sebagai sebuah perangkat aturan untuk mengefesiensikan sebuah pengujian yang luas terhadap bidang pengujian yang lebih sederhana dengan tetap memperhatikan maksudnya. Tujuan dari target dilakukan pengklarifikasian pada sekelompok yang lebih rumit, setelah kemudian keputusan dari pohon bertindak sebagai panduan dengan lebih jelas memperhitungkan probabilitas hampir semua record relative terhadap bagian tertentu. Pohon Keputusan dapat dijabarkan berdasarkan algoritma yang diterapkan agar menghasilkan proses terkumpulnya data dimana setiap ruangan data tidak dapat dipahami (Eska, 2018)



## Gambar 2.4 Model Pohon Keputusan

**Sumber:** (Hamidah, 2020)

### 2.1.1.5 Algoritma C4.5

Algoritma dengan cara penggunaan dipermudah dalam membentuk suatu system pendukung keputusan dengan berbentuk pohon yang dapat membantu dalam mengambil keputusan. Algoritma C4.5 dapat dimaksud juga dengan suatu perubahan yang telah berhasil dilaksanakan pada algoritma ID3 hingga tercapai algoritma C4.5, maka algoritma C4.5 dapat dikatakan juga sebagai algoritma yang berasal dari algoritma ID3. Dalam menyelesaikan masalah yang ada algoritma C4.5 memberikan solusi pada kehilangan informasi dan memberikan solusi saat terjadinya perpecahan data. Alur pembentukan bidang yang terstruktur sering dilakukan penggunaan algoritma C4.5 agar terbentuk pohon keputusan (Fikri & Verina, 2020).

Algoritma C4.5 dirancang sebagai mekanisme dengan karakteristik pertumbuhan keputusan dari pohon. Konsep pemikiran pohon keputusan dimulai pada akar yang dapat dikatakan sebagai atribut yang paling tinggi yang kemudian berdasarkan percabangan yang lebih sederhana dapat diakhiri dengan atribut daun yang memiliki atribut paling rendah. Atribut yang paling tinggi dapat dikatakan juga sebagai *node* dimana atribut tersebut adalah atribut yang memiliki nilai lebih tinggi dari pada atribut yang ada

## **2.1.2 Teori Khusus**

Penulis berupaya menjelaskan mengenai teori yang berhubungan dengan sejumlah kenyataan yang bersifat khusus, dengan berusaha menjelaskan hubungan antara satu sama lain antara teori dan permasalahan yang akan diteliti.

### **2.1.2.1 Bonus**

Bonus ialah semua bentuk dari sebuah benda atau barang yang dapat diberikan yang dapat berwujud maupun tidak berwujud yang harus diperhitungkan dan diterima setiap para pekerja yang seharusnya mendapatkan. Sebuah skema yang telah diperhitungkan dengan benar akan menghasilkan kelegaan dari para pekerja sehingga bisa mendapatkan hasil yang diinginkan setiap perbulannya rewa (Dhamayanti, 2018)

### **2.1.2.2 Bonus Tahunan**

Segala sesuatu yang dapat diterima baik berupa fisik hingga non fisik yang kemudian harus dilakukan perhitungan dan diserahkan setiap tahun kepada karyawan yang berhak menerimanya agar karyawan dapat merasakan kepuasan dan meningkatkan kinerja karyawan untuk memperoleh keuntungan pada perusahaan atau instansi tersebut. (Afriany & Aisyah, 2019).

### **2.1.2.3 Karyawan**

Factor pendukung dari suatu perusahaan atau instansi adalah karyawan, karena dengan kehadiran karyawan yang sesuai dengan kriteria yang telah ditentukan berdasarkan standar dari perusahaan dan instansi tersebut dengan melakukan kualifikasi terhadap setiap karyawan akan bekerja maka suatu

perusahaan atau instansi tersebut dapat terus berkembang (Afriany & Aisyah, 2019).

#### **2.1.2.4 Kinerja Karyawan**

Kuantitas maupun kualitas pekerjaan yang dihasilkan oleh karyawan seperti prosedur kerja, system kerja, operasional kerja, dan kriteria yang telah ditentukan pada suatu perusahaan atau instansi merupakan kinerja karyawan. Kriteria yang dapat digunakan dalam mengukur kinerja karyawan adalah hasil dari kuantitas yang telah dikerjakan, hasil dari kualitas yang telah dikerjakan, menyelesaikan pekerjaan sesuai dengan waktu yang telah ditentukan, kedisiplinan terhadap kehadiran, dan mampu bekerja sama (Fransisca Andreani & Tanto Wijaya, 2019).

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

Sebuah perkumpulan yang solid yang dapat dikatakan juga sebagai gudang dari berbagai tingkat java yang digunakan untuk meningkatkan sekumplan informasi dengan mesin yang dapat dipergunakan sebagai alat studi algoritma seperti data mining dapat disebut juga dengan WEKA.

Dengan sebuah Gudang dari berbagai tingkat yang terkomputerisasi luas dengan penerapan secara online sangat mempermudah pengoperasian dalam penerapan mesin studi terhadap informasi yang berasal dari pengoperasian yang didasarkan pada komputerisasi (Azwanti, 2018).



**Gambar 2.5** Tampilan Utama WEKA

## 2.2. Penelitian Terdahulu

Peneliti juga menemukan beberapa hasil penelitian terdahulu yang berkaitan dengan menganalisa menggunakan algoritma C4.5 yang dirangkum yaitu:

1. Dari penelitian “Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT. Capella Dinamik Nusantara Cabang Muka Kuning yang dilakukan oleh (Azwanti, 2018) didapatkan informasi bahwa jenis algoritma dari C4.5 bermanfaat dalam kegiatan atau proses melakukan klasifikasi kemudian juga dapat digunakan untuk tujuan prediksi penjualan dari motor. Pada tahap atau proses pengujian dengan aplikasi WEKA ternyata juga memberikan kemudahan bagi peneliti.
2. Dari penelitian “Penerapan Algoritma C4.5 Untuk Memprediksi Penjualan Sayur Pada PT Bumi Bhakti Bareleng” dari (Harman, 2022)

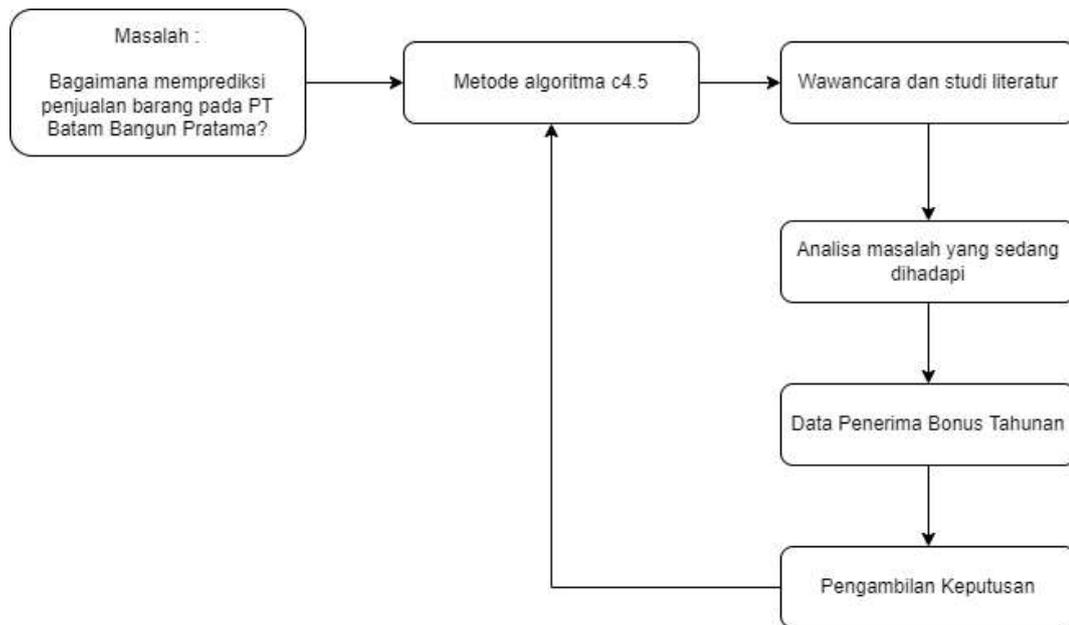
didapati penelitian ini bermanfaat untuk mempermudah bagian penyedia stok sayur pada PT Bhakti Bumi Bareleng dalam melakukan perencanaan penyediaan stok sayur serta memberitahu pihak perusahaan tentang produk-produk paling banyak dibeli oleh konsumen.

3. Sebuah observasi berjudul “Prediksi Profit Perusahaan Menggunakan Algoritma C4.5” dari (Elisa, 2018) ditemui hasil berikut yakni penerapan jenis algoritma C4.5 telah dapat dibuktikan mampu merangkai sebuah sistem yang menyokong pengambilan putusan dalam membuat prediksi-prediksi peluang sebuah entitas bisnis berdasarkan kriteria yang telah ditetapkan. Penerapan jenis algoritma C4.5 juga bernilai memiliki efektifitas yang cukup baik.
4. Dari penelitian “*Student Performance Analysis Using C4.5 Algorithm to Optimize*” yang dilakukan oleh (Hilda Amalia, Yunita, Ari Puspitasari, & Ade Fitria Lestari, 2020) dalam penelitian ini dapat disimpulkan bahwa metode seleksi optimasi tersebut dapat berjalan dengan baik sehingga dapat meningkatkan nilai akurasi yang lebih baik dari akurasi nilai yang dihasilkan hanya dengan menggunakan algoritma C4.5 metode data mining.
5. Dari penelitian “*Implementation of C4.5 Algorithm to Assist in The Selection of Floor Construction*” yang dilakukan oleh (Aditya Roval Lendra1 & Diky Firdaus, 2020) didapatkan informasi bahwa hasil penelitian yang dilakukan dapat menentukan proyek mana yang dapat diambil berdasarkan atribut yang telah ditentukan.

6. Dari penelitian “Analisis Algoritma C.45 Untuk Menentukan Faktor Pembelian Sepeda Bekas Pada Toko Sepeda Batam” yang dilakukan oleh (Harman, 2021) didapatkan informasi bahwa penggunaan jenis algoritma dari C4.5 diharapkan dapat membantu perusahaan dalam mempelajari keputusan pembeli saat membeli sepeda bekas.
7. Penelitian yang berjudul “*Graduation Prediction System on Students Using C4.5 Algorithm*” yang dilakukan oleh (Donny Kurniawan, Anthony Anggrawan & Hairani, 2020) menghasilkan bahwa implementasi algoritma C4.5 sangat layak untuk digunakan sebagai prediksi pelajar untuk kelulusan.
8. Dari penelitian ”Data Mining Menggunakan Algoritma C4.5 Untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Di Kota Batam yang dilakukan oleh (Yulia & Putri, 2019) didapatkan informasi bahwa hasil prediksi mengenai kinerja Dosen dengan menggunakan dasar nilai tertinggi yang terdapat pada gain. Kemudian struktur decision tree yang dibuat dengan model perhitungan secara manual dengan menggunakan jenis algoritma dari C4.5 menghasilkan pola berbentuk aturan yang dapat menjelaskan proses yang terjadi dalam tahapan prediksi yang berhubungan dengan variabel dari penelitian yang dilakukan. Penggunaan atau cara data mining dengan menggunakan jenis algoritma C4.5 secara manual ternyata terbukti memiliki hasil berupa nilai akurasi yang sama dengan penggunaan aplikasi WEKA.

### 2.3. Kerangka Pemikiran

Berikut adalah kerangka pemikiran dalam penelitian ini, yang antara lain dapat digambarkan sebagai berikut:



**Gambar 2.6** Kerangka Berpikir  
Sumber: (Peneliti, 2022)