

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori Dasar**

Penulisnya memaparkan berbagai teorinya yang dijadikan sebagai acuan yang mendasar didalam melakukan penelitiannya dengan menjabarkan definisi serta istilah yang terkait dengan judul penelitiannya.

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

Penulisnya menganalisis berbagai teorinya yang umum untuk melakukan penelitian yang mencakup definisi dan istilahnya yang umum.

##### **2.1.1.1 Sistem Pendukung Keputusan (SPK)**

Sebagian dari sistem informasinya terkomputerisasi seperti sistem yang basisnya pengetahuan dan manajemen pengetahuannya, dan di gunakan didalam memotivasi penentuan keputusannya sebuah perusahaan. Pada dasarnya, DSS adalah evolusi dari sistem informasinya suatu manajemen yang di komputerasikan serta di rancang untuk saling melakukan interaksi sesama pengguna. Sistem ini mencakup sejumlah komponennya yang utama, yakni database, model base, dan dialog.

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

KDD ialah langkah didalam memperoleh informasi melalui penggunaan data dari database atau data yang disimpan. Setelah informasi ini ditemukan,

digunakan sebagai basis pengetahuan dalam penentuan keputusannya. Berikut tahapannya didalam KDD, yakni:

1. *Data Selection*

Fasenya ini berjalan diawal prosesi KDD. Artinya, dilakukan dengan mengumpulkan informasinya dengan cara pemilihan datanya yang di jadikan didalam sumber datanya yang diakhirnya diproses ditahap data mining.

2. *Pre-processing atau Cleaning*

Tahapan ini di lakukan didalam membuang dan memvalidasi sejumlah duplikasinya pada datanya yang tidak memenuhi, sebab prosesnya didalam entry datanya haruslah mencukupi keperluan penelitiannya.

3. *Transformation*

Langkah didalam mengubah serta mengadaptasi model menyimpan database.

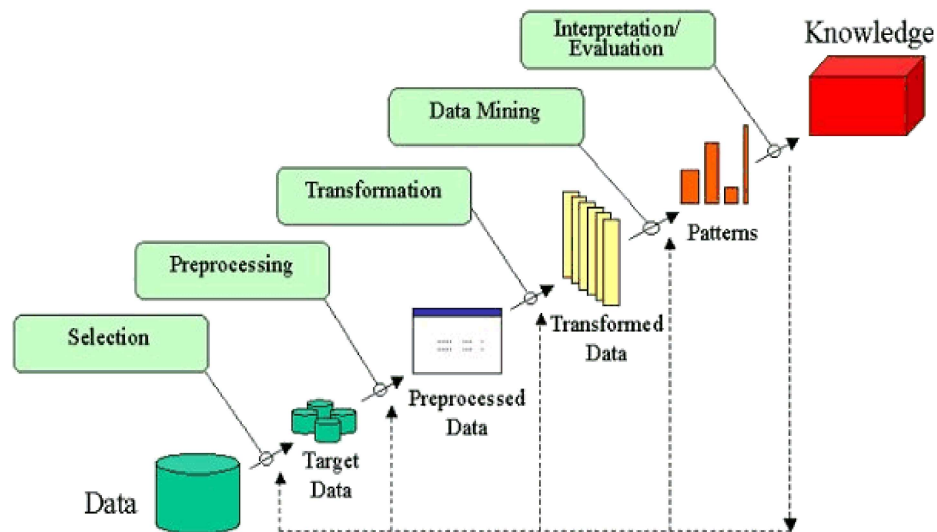
4. *Data Mining*

Suatu langkah didalam menyediakan caranya untuk menemukan serta menyerupai pola peraturan didalam memperoleh informasinya didalam keputusannya berdasar tujuan penelitiannya.

5. *Interpretation*

Hal ini membuat hasilnya sebuah model peraturan data mining mudah di pahami. Secara khusus, ini menampilkan informasi yang bertentangan dengan hipotesis penelitian.

Langkah didalam KDD yakni:



*Gambar 2.1 Proses KDD*

### 2.1.1.3 Data Mining

Langkah didalam menemukan informasinya yang sebelumnya tak di ketahui dari kumpulan datanya yang besar. Data mining ialah rangkaian prosesi yang menggunakan 1 ataupun lebih tekniknya didalam mempelajari komputer didalam menganalisa serta mengekstrakkan pengetahuannya dengan otomatis, ataupun rangkaian prosesi yang menemukan penilaian dari sekumpulan datanya sebagai pengetahuannya. Hanya mereka yang tak di definisikan dengan manual. Data mining di gunakan menggambarkan menemukan pengetahuannya didalam database. Data mining ialah prosesi mengekstraksikan serta mengidentifikasi informasi dan pengetahuannya yang bermanfaat dengan menerapkan teknik statistika, matematika, kecerdasan buatanya, serta mempelajari mesin. Dari database yang besar. Di lihat dari definisinya mining artinya mengekstraksi dari data stack. Penumpukan datanya di sini ialah datanya yang sudah di kumpulkan sebelumnya. Tujuannya dari data mining ialah didalam mendapatkan sekumpulan

informasinya ataupun pengetahuannya yang dapat disimpan selama data mining serta digunakan oleh komunitas dan organisasi itu sendiri (Yulia & Azwanti, 2018).

#### **2.1.1.4 Pengelompokkan Data Mining**

Apabila pembagiannya di dasarkan berdasar tugasnya serta langkahnya yang memiliki basisnya didalam prosesnya yang di lakukan didalam data mining, bisa di bagi menjadi sejumlah tahapan berikut:

1. Deskripsi atau *Description*

Fase ini dimulai dengan seorang analisanya yang mencoba didalam melakukan klasifikasi atas jenisnya yang terkandung didalam datanya serta sumber informasinya. Hal ini bisa di jelaskan apabila staff yang bertanggung jawab untuk menghitung aktivitas didalam memungut suaranya didalam mengkampanyekan pemilu tak diketahui ataupun memperoleh informasinya yang lebih jelas terkait keterampilannya serta pengalaman para kandidatnya yang terdaftar dalam pemilu. Oleh sebab itu, penjabaran permasalahannya ialah pengupayaan didalam memerikan penjabatannya terkait kemampuannya pihak yang ada didalam sistem pemilihannya.

2. Estimasi atau *Estimation*

Fase ini aktivitasnya yang di lakukan dari estimasi berfungsi sama dengan pengklasifikasiannya, namun berbeda jika proses estimasi atau variabel objektif yang terdapat pada fase ini mengandung angka. nilai. data. Pola yang dibuat bersumber dari kumpulan datanya yang berisikan angka

ataupun variabelnya yang ada di dalamnya yang di gunakan sebagai bahan penetapan nilainya didalam proses memprediksikan. Lalu dilanjutkan dengan memberikan estimasinya yang diturunkan dari nilainya yang ada divariabel yang diturunkan dari nilai prediksinya. Hal ini bisa di jelaskan dengan nilainya yang termasuk didalam IPK ataupun nilai rata-rata (S.1) mahasiswa yang sedang mengikuti program pendidikan di universitas tersebut. Estimasi dapat ditampilkan dengan menggunakan nilai IPK atau Indeks Prestasi Mahasiswa selama proses pengajaran setiap semester.

### 3. Prediksi atau *Prediction*

Proses atau fase peramalan juga hampir mirip dengan proses dan fase klasifikasi dan juga selama proses estimasi, tetapi tidak termasuk apakah ramalan nilai yang ada ditujukan untuk masa depan.

### 4. Klasifikasi atau *Classification*

Aktivitas yang dikategorikan menggunakan beberapa target untuk variabel berdasarkan kategori yang ada. Inilah ketika mengklasifikasikan pendapatan karyawan suatu perusahaan, pendapatan karyawan dibagi menjadi tiga kategori: karyawan berpenghasilan sangat tinggi, karyawan berpenghasilan menengah atau berpenghasilan normal, sangat rendah.

### 5. Pengklusteran atau *Clustering*

Langkah yang didalamnya ada pengelompokkan ialah kegiatannya yang melakukan pengelompokkan hasil pengumpulan datanya, serta datanya di kelompokkan kembali berdasar datanya yang menunjukkan kesamaan ataupun pola kesamaan. Kelompok yang bersangkutan memiliki hasil

pencatatan pada satu kelompok yang memiliki pola yang sama atau serupa, dan kelompok yang lain tidak memiliki pola yang sama atau serupa pada hasil pencatatan yang ada. Satu atau lebih proses dalam sebuah cluster berbeda dari satu atau lebih proses dalam klasifikasi karena cluster tak mempunyai target. Algoritma komputasi yang di gunakan didalam clustering berfokus pada mengelompokkan datanya yang tersedia berdasar kesamaannya yang berfokus model, sehingga pekerjaan yang dilakukan pada tahap dan proses clustering adalah estimasi, prediksi, dan tidak melakukan proses yang berhubungan dengan klasifikasi dan untuk evaluasi. Kemiripan atau kemiripan data yang disimpan dalam kelompok mencapai maksimum atau sebaliknya. Jika kemiripan ataupun kesamaan datanya yang tersedia sangatlah rendah, makanya hasil dari penilaian ini cukup kecil.

#### 6. Asosiasi atau *Association*

Asosiasi tersebut melalui proses memperoleh informasi lengkap atau pendukung yang terkandung dalam suatu item data.

#### **2.1.1.5 Pohon Keputusan (Decision Tree)**

Pohon keputusan menerapkan representasi pohon dimana tiap node mewakili atributnya, cabang mewakili nilai atributnya, serta daunnya mewakili kelasnya. Node teratas pada pohon keputusannya di sebut root. Pohon keputusannya ialah metode kategorisasi yang umumnya di gunakan. Selain cukup cepat didalam penyetaan, hasilnya dari modelnya yang di bangun juga mudah di pahami.

Data pohon keputusannya umumnya di representasikan berbentuk tabel atributnya dan catatannya. Atributnya mewakili parameternya yang di sebut karakteristik bangunan pohon. Keuntungan utamanya menerapkan pohon keputusan ialah anda dapat memecah proses keputusan yang awalnya rumit menjadi lebih ringan. Hal ini memungkinkan penentuan keputusannya menginterpretasikan solusinya untuk masalah tersebut. Jika Anda memutuskan untuk bermain tenis, kriterianya yang perlu dipertimbangkan ialah cuacanya, anginnya, dan suhunya. Sebuah atribut disebut atribut hasil, yang mewakili data solusi untuk setiap item data.

Ada 3 jenis node dalam pohon keputusannya, yakni:

1. *Root node* ialah node teratas, tak adanya penginputan pada nodenya serta mungkin tak mempunyai outputnya ataupun mempunyai lebih dari satu outputnya.
2. *Internal Node* ialah node cabangnya, dengan nodenya ini hanyalah mempunyai 1 inputnya dan mempunyai setidaknya 2 output.
3. *Leaf Node* ataupun Terminal Node ialah nodenya yang terakhir, hanyalah ada 1 input serta tak adanya outputnya di node ini. (Amalia, 2018).

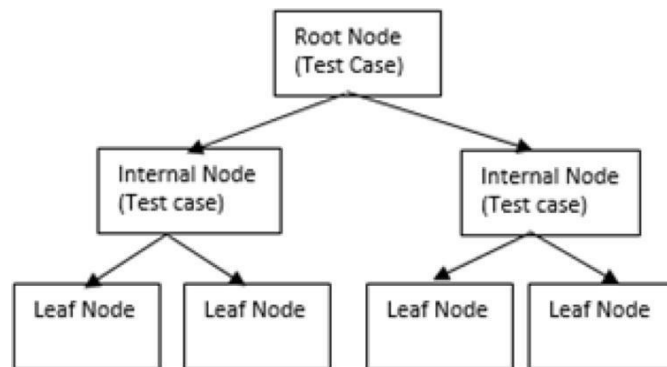
Selanjutnya berikut penjelasan mengenai kelebihan dan kekurangannya dari pohon keputusannya, yakni:

1. Kelebihan
  - a. Dalam penggunaan pohon keputusannya, ketika ada cakupan data kompleks yang telah di modifikasi supaya lebih memudahkan disaat di gunakan nantinya dianggap keputusan kebijakannya.

- b. Bila membuang metodenya didalam perhitungannya yang tak diperlukan, sehingga hanyalah berkriteria yang memenuhi kebutuhannya didalam pengujian yang digunakan dalam proses pengujian.
  - c. Jika anda memilih integritas atau deskripsi data saat menggunakan pohon keputusan, hasilnya dari pemilihannya terintegritas ataupun keakurasian datanya ialah perbedaan peringkat datanya berdasar kriterianya yang lain.
  - d. Membantu menghindarkan permasalahan lainnya dengan memakai beberapa keterbatasan jenis kelompoknya ataupun kategorinya tanpa melakukan pengorbanan kualitasnya dari hasil keputusannya yang terakhir.
2. Kekurangan
- a. Disaat memakai pohon keputusannya, sejumlah besar hasil keputusannya untuk dihasilkan serta kriterianya yang di gunakan akan meningkatkan waktunya yang di butuhkan memperoleh keputusannya serta perlu sistemnya dengan dukungan memorinya yang tinggi.
  - b. Pohon keputusannya mempunyai jumlah sekumpulan kesalahannya yang ada ditiap bagian strukturnya yang ada dipohon keputusannya.
  - c. Sulit didalam mendirikan struktur pohon keputusannya yang memerikan hasilnya yang terbaik.



- d. Disaat pohon keputusannya digunakan, hasilnya ataupun kualitasnya yang di hasilkan sangatlah di tentukan oleh modelnya ataupun strukturnya yang diimplementasikan.



*Gambar 2.2 Model Pohon Keputusan*

#### **2.1.1.6 Algoritma C4.5**

Algoritma C4.5 membantu didalam mengklasifikasikan datanya ke dalam angkanya ataupun kategorinya. Sesudah tahapan prosesnya pengklasifikasiannya selesai, makanya akan dihasilkan aturan pola yang berbeda yang bisa di gunakan didalam melangsungkan tahapan prosesnya didalam memprediksikan didalam memakai nilai prediksinya yang di peroleh dengan merekam datanya yang lama ke datanya yang terbaru.

Algoritma C4.5 di gunakan didalam mendirikan pohon keputusannya. Sejumlah pengembangannya yang di lakukan oleh di C4.5 termasuk melebihi nilai yang hilang, melebihi data yang berkelanjutan, dan pemangkasan. (Tukino, 2019). Menurut (Hendra & Harman, 2020) Ada sejumlah langkah didalam membangun pohon keputusannya pada algoritma C4.5:

1. Persiapkan data pelatihannya. Datanya latih umumnya di ambil dari data historisnya yang sudah dibangkitkan sebelum nya ataupun data historis yang bisa mereferensikan data sebelum nya serta dikategorikan kedalam kelas tertentu.
2. Perhitungan akar pohonnya. Akar diekstraksi dari atributnya yang di pilih memakai metode penghitungan perolehan *gain* untuk tiap atributnya dengan perolehan *gain* yang paling tinggi. Dengan kata lain, ini adalah akar pertama dari pohon keputusan. Langkah pertama adalah menghitung perolehan *entropy* sebelum menghitung perolehan *gain*.

Perhitungan I untuk menghitung *entropy*:

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

**Rumus 2.1 Perhitungan Entropy**

Keterangan:

1. Simbol *S* ialah himpunan
2. Simbol *A* ialah keterangannya yang di gunakan.
3. Simbol *n* ialah banyaknya partisi yang ada diketerangan bagian *S*  $p_i$  ialah proporsinya dari  $S_i$  kepada *S*.
4. Dalam perhiutungan nilai *gain* menggunakan persamaan II:

$$\mathbf{Gain (A) = Entropi (S) - \sum_i^n \frac{|S_i|}{|S|} x Entropi (S_i)}$$

**Rumus 2.2 Perhitungan Gain**

Keterangan:

1. Simbol  $S$  ialah himpunan
2. Simbol  $A$  ialah keterangan yang di gunakan.
3. Simbol  $n$  ialah jumlah partisi yang ada diketerangan bagian  $A$ .
4. Simbol  $|S_i|$  ialah banyaknya dari kasusnya yang ada didalam partisinya yang ke- $i$ .
5. Simbol  $|S|$  ialah banyaknya dari kasusnya yang ada didalam  $S$ .

### 2.1.2 Teori Khusus

Penulisnya menjabarkan sejumlah teori khususnya untuk melakukan penelitiannya mencakup istilah serta definisi tertentu.

#### 2.1.2.1 Prediksi

Prediksi ialah aktivitas memprakirakan ataupun memprediksikan hal yang hendak terjadi dimasa mendatang didalam waktunya yang singkat. Prognosis ialah keadaan yang kemungkinan terjadi di masa mendatang.

##### 1. Jenis Prediksi

Tergantung pada jenis prediksi itu dapat dibagi menjadi beberapa jenis.

Menurut perencanaan pengoperasiannya dimasa mendatang, ramalan terbagi kedalam 3 jenis, yakni:

- a. Prediksi ekonomi, menggambarkan siklusnya sebuah perusahaan dengan memprediksikan tingkatan inflasinya, adanya dana, kebutuhan dananya untuk membangun rumah, dan lainnya.

- b. Prediksi teknologi, membahas majunya teknologi yang bisa membawa dan menarik produknya yang baru ke pasar yang memerlukan pabriknya serta peralatannya yang baru.
- c. Prediksi permintaan, ialah prakiraan permintaannya didalam layanannya ataupun produknya perusahaannya.

## 2. Metode Prediksi

Analisi *Time Series* ialah hubungannya diantara variabelnya yang Anda cari dengan variabelnya yang bisa mereka pengaruhi, terkait dengan waktunya misalnya mingguan, bulanan, triwulanan, triwulanan, setengah tahunan atau tahunan.

Didalam analisa *time series*, variabel yang di cari ialah waktu.

Metode prediksinya mencakup:

- a. Metode *Smoothing* adalah beberapa tipe ramalannya yang berjangka pendek, mencakup rencana keuangannya, rencana persediaannya. Tujuan menerapkan metodenya ini ialah demi mendeteksi ketidakteraturan didata masa lalunya, misalnya musiman.
- b. Metode *Box Jenkins* ialah pengaturan waktunya yang digunakan untuk prediksi jangka pendek dengan menggunakan model matematika.
- c. Metode proyeksi tren regresi ialah metodenya yang di gunakan didalam jangkanya yang pendek maupun panjang. Metodenya

ini sebagai rumusan garis trend didalam persamaannya yang matematis.

#### **2.1.2.2 Penjualan**

Definisi penjualan berarti mikro berupa organisasinya yang memprakirakan kebutuhannya para pelanggan serta mengatur arus barangnya dan jasanya didalam mencukupi kebutuhan produsennya serta konsumennya sedangkan mikro bersifat sosial, yaitu mengatur kegiatan untuk mencapai tujuan kita. Proses membimbing arus barangnya dan jasanya dalam perekonomian dari produsennya ke konsumennya, seefektif mungkin didalam mencocokkan penawarannya serta permintaannya, dan meraih tujuan sosialnya.

Penjualan ialah aktivitas yang memiliki tujuannya didalam memenuhi kebutuhannya dan keinginannya pelanggan dengan cara menukarkan produk dan prospek di dalam perusahaan.

#### **2.1.2.3 Kualitas Produk**

Menurut (Umboh, Tumbel, & Soepeno, 2015) pengertian kualitas produk perlu dijadikan perhatiannya produsen atau perusahaannya, sebab kualitasnya dari produk yang berkaitan dengan kepuasan konsumennya yang dijadikan tujuannya dari aktivitas pemasarannya. Kotler (2005: 49) menjabarkan kualitas produk sebagai karakteristik dan ciri khusus umum dari sebuah produknya ataupun jasanya didalam berkemampuan didalam memuaskan kebutuhannya yang dinyatakan / tersirat.

## 2.2 Penelitian Terdahulu

Penelitinya juga telah menemukan sejumlah temuan sebelumnya berkaitan analisis menggunakan algoritma C4.5 yang disimpulkan yaitu:

1. Dengan judul “Penerapan Algoritma C4.5 Untuk Penentuan Penerima Beasiswa Mahasiswa” yang di lakukan oleh (Hasmin & Aisa, 2019) penelitian ini menunjukkan bahwa mengidentifikasi memakai teknik data mining terutama algoritma C4.5 dapat sangat membantu didalam pengambilan keputusan.
2. Dengan judul “Penerapan Algoritma C4.5 Untuk Memprediksi Besarnya Penggunaan Listrik Rumah Tangga di Kota Batam” yang dilakukan oleh (Yulia & Azwanti, 2018) didalam penelitiannya ini, hasil perhitungan dengan memakai algoritma C4.5 memperhatikan lima atribut yang digunakan yaitu pendidikan, jumlah keluarga, dan nilai *gain* tertinggi dari semua produk elektronik menunjukkan bahwa konsumsi dapat diprediksi dari jumlah pengguna, lama tinggal di rumah, luas bangunan rumah.
3. Dengan judul “Penerapan Data Mining Untuk Prediksi Penjualan Alat Medis Menggunakan Algoritma C4.5 PT. Murni Indah Sentosa” yang dilakukan oleh (Fikri & Verina, 2020) penelitian ini menunjukkan bahwa algoritma *data mining* C4.5 ini, PT. Murti Indah Sentosa digunakan untuk memberikan kemudahan dalam mengolah data penjualan alat kesehatan dengan cara cepat mendokumentasikan data penjualan dan menyimpannya dalam database.

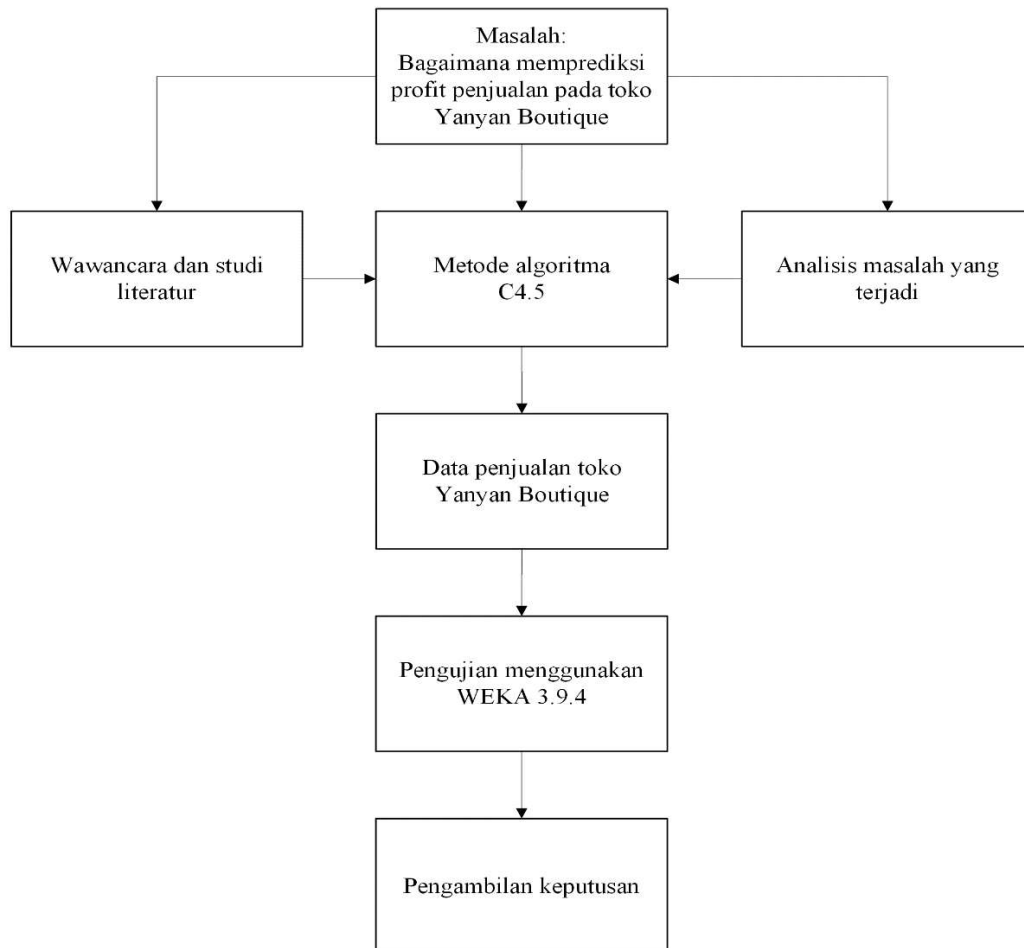
4. Dengan judul “Data Mining Algoritma C4.5 Untuk Memprediksi Penjualan Baterai Di PT Varta Microbattery Indonesia” yang dilakukan oleh (Wijaya & Fauzi, 2020) penelitian ini menunjukkan bahwa Pengujiannya dengan memakai software RapidMiner begitu efektif di karenakan pohon keputusannya yang di dapat hasilnya serupa dengan perhitungannya secara manual. Oleh sebab itu memakai RapidMiner begitu membantu didalam pengujiannya yang di lakukan.
5. Dengan judul “Penerapan Data Mining Untuk Prediksi Penjualan Readymix Menggunakan Metode Algoritma C4.5 Pada PT Remicon Widyaprima” yang di lakukan oleh (Hendra & Harman, 2020) penelitiannya ini menggunakan perhitungan pengklasifikasian untuk referensi data transaksi penjualan PT Remicon Widyaprima atau sebagai panduan untuk membangun pohon keputusan disaat memakai teknik *data mining* dengan algoritma C4.5 didalam membuat prakiraan penjualan ready mix.
6. Dengan judul “Penerapan Algoritma C4.5 didalam Memprediksikan Keuntungan Pada PT SMOE Indonesia” yang dilakukan oleh (Tukino, 2019) penelitiannya ini menggunakan algoritma C4.5 sebagai metode klasifikasi dan mempertimbangkan perolehan *gain* yang paling tinggi, hingga menguntungkan PTSMOE Indonesia.
7. Dengan judul “Analisis Kualitas Produk, Brand Image Dan Life Style Terhadap Keputusan Pembelian Pakaian Wanita Di Mississippi Manado Town Square” yang di lakukan oleh (Umboh, Tumbel, & Soepeno,

- 2015) penelitian ini menunjukkan bahwa temuan tersebut dapat digunakan untuk mengembangkan ilmu manajemen pemasaran.
8. Dengan judul “Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa” yang dilakukan oleh (Swatisna, 2013) penelitiannya ini menunjukkan Algoritma Decision Tree C4.5 akurat diterapkan didalam menentukan kesesuaiannya jurusan mahasiswa.
  9. Dengan judul “Analisis Kepuasan Pelanggan Dengan Algoritma C4.5 Pada PT Wook Global Technology” yang dilakukan oleh (Amalia, 2018) penelitian ini menunjukkan bahwa penelitian ini dilakukan pemodelan memakai datanya yang di olah berdasar tahapan KDD.
  10. Dengan judul “Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus PT. Indomarco Palembang” yang dilakukan oleh (Sutrisno, Afriyudi, & Widiyanto, 2013) penelitian ini menunjukkan bahwa penerapan *data mining* memakai aplikasinya yang di bangun bisa membantu PT. Indomarco sebagai contoh proses penentuan keputusannya suatu perusahaannya untuk memperoleh sampel penjualan produknya.



### 2.3 Kerangka Pemikiran

Didalam kerangka pemikirannya ini ialah langkah yang di lakukan didalam menyelesaikan permasalahan penelitiannya yang di buat, yakni:



**Gambar 2.3** Kerangka Pemikiran

**Sumber:** (Penulis,2022)