

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori**

Pada bagian ini, akan diuraikan makna dan definisi yang relevan dengan topik penelitian ini, serta akan dibahas beberapa teori-teori dasar yang dijelaskan sebagai berikut:

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

Dalam bagian ini, akan dijelaskan beberapa konsep teoritis yang menjadi dasar pemahaman bagi penulis, yang dianggap penting untuk memperkuat argumentasi dan kesimpulan yang ingin diperoleh melalui penelitian ini, dan teori tersebut bersifat umum dan luas.

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

KDD, yang merupakan kependekan dari *Knowledge Discovery in Database*, dapat dijelaskan sebagai salah satu metode untuk memperoleh informasi atau pengenalan dengan menggunakan bukti-bukti yang dianalisis dari basis data atau bukti yang disimpan pada suatu tempat (Panggabean et al., 2020). Informasi yang ditemukan ini kemudian digunakan sebagai dasar pengetahuan untuk mendukung pengambilan keputusan yang berkaitan dengan permasalahan yang ada. Berikut adalah tahapan-tahapan dalam *Knowledge Discovery in Database* dalam penelitian yang dilakukan (Asroni et al., 2018):

1. Pemilihan Data

Langkah awal yang penting adalah mengumpulkan semua data atau bukti yang diperlukan melalui seleksi tertentu, sehingga membentuk sumber data yang dapat diandalkan sebelum melanjutkan ke proses *data mining*.

## 2. Pra-pemrosesan atau Pembersihan

Langkah berikutnya adalah tindakan preventif yang bertujuan untuk menghilangkan data ganda dengan melakukan pemeriksaan yang relevan sesuai dengan kepentingan yang mungkin muncul selama proses analisis data. Hal ini terjadi karena bukti-bukti yang dapat bertambah seiring berjalannya penelitian.

## 3. Transformasi

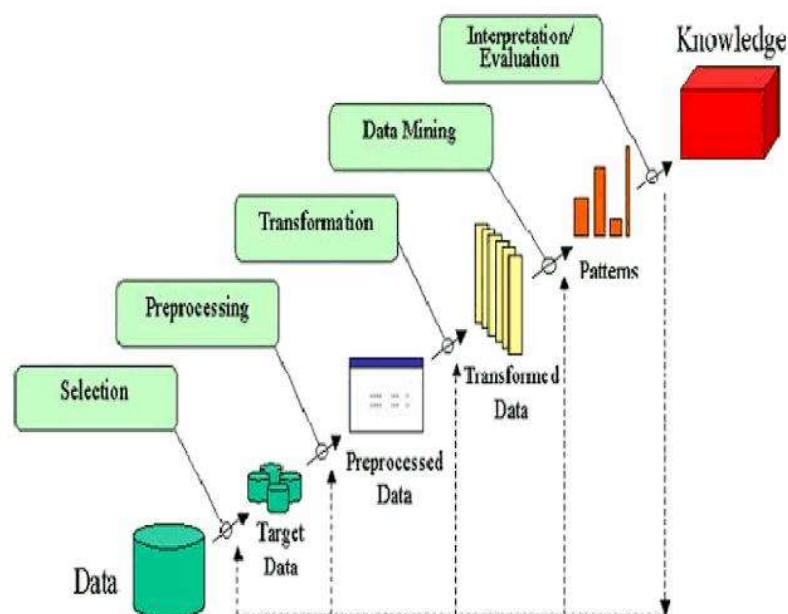
Kemudian, proses ini melibatkan perubahan dan penyesuaian pola penyimpanan data dasar.

## 4. Pencarian Data (*Data Mining*)

Tahap yang sangat penting dalam rangkaian langkah sebelumnya adalah menentukan upaya dalam proses pencarian data dan pembentukan aturan yang menggambarkan bukti-bukti yang ada. Ini akan membimbing pengambilan keputusan sesuai dengan identifikasi masalah yang harus dipecahkan.

## 5. Interpretasi

Tahap akhir dari seluruh proses adalah memberikan kesan atau perkiraan dari tren atau fenomena yang ditemukan melalui *data mining*, sehingga dapat dengan mudah dianalisis berdasarkan pemahaman yang sederhana, terutama bukti yang relevan dengan tujuan penelitian.



**Gambar 2.1** Proses *Knowledge in Database*

**Sumber:** (Nasari & Darma, 2015)

### 2.1.1.2 *Data mining*

*Data Mining* adalah proses yang kadang disebut sebagai penemuan informasi yang awalnya tidak diketahui, yang berasal dari kumpulan data yang besar, dapat dianggap sebagai suatu teknik yang merujuk pada bidang ilmu yang signifikan, dan mampu mengungkapkan pola dalam kumpulan data yang luas (Siahaan & Fauzi, 2023). Definisi lain dari *data mining* adalah suatu rangkaian metode yang menggunakan teknik studi komputerisasi dalam menganalisis dan menyajikan informasi secara otomatis, yang juga dapat didefinisikan sebagai suatu rangkaian metode sistematis yang digunakan untuk mendapatkan informasi yang masih kurang lengkap tanpa perlu mencarinya secara manual dalam ilmu yang relevan (Irawan Yudi, 2019).

### 2.1.1.3 Pengelompokan *Data Mining*

*Data Mining* dalam konteks ini adalah serangkaian tahapan yang dapat diuraikan secara lebih rinci, yang bertujuan untuk memperjelas langkah-langkah yang harus diikuti guna mencapai interpretasi yang kuat dalam konteks penjualan makanan kuaci. Tahap-tahapnya menurut penelitian (Mardi Yuli, 2017) adalah sebagai berikut:

#### 1. Deskripsi (*Description*)

Tahap awal, yang bertindak sebagai fondasi untuk seluruh proses yang akan datang, melibatkan analisis untuk mendeskripsikan ciri-ciri khusus dari data yang terkait dengan penjualan makanan kuaci. Proses ini mirip dengan mengumpulkan suara dalam pemilihan umum tanpa terpengaruh oleh informasi yang tidak relevan. Dalam konteks penjualan kuaci, ini berarti memaparkan informasi terkait data dalam pola yang terstandarisasi.

#### 2. Estimasi (*Estimation*)

Langkah berikutnya dari deskripsi adalah mengestimasi, atau dengan kata lain, mengklasifikasikan data satu dengan yang lain. Ini menjadi penting ketika data memiliki nilai numerik. Hasil estimasi ini diwakili dalam bentuk angka, yang membantu dalam memprediksi hasil. Misalnya, ini bisa menggambarkan sejauh mana pemahaman konsumen tentang kuaci berdasarkan pengelompokan hasil seperti pada nilai rapor siswa.

#### 3. Prediksi (*Prediction*)

Tahap selanjutnya adalah memprediksi hasil dari data yang telah dideskripsikan dan diestimasi sebelumnya. Fokus utama pada tahap ini adalah memahami apa yang akan terjadi di masa depan, sehingga menjadi alat dalam pengambilan keputusan. Dalam konteks penjualan kuaci, ini bisa berarti:

- a. Mencari peluang kenaikan atau penurunan penjualan kuaci di bulan atau tahun-tahun mendatang.
- b. Meramalkan peluang kenaikan atau penurunan penjualan kuaci dengan tipe produk tertentu di masa depan.
- c. Menduga kemungkinan naik atau turunnya permintaan untuk bahan bangunan yang terkait dengan listrik dan air bersih dalam bulan atau tahun-tahun mendatang.

#### 4. Klasifikasi (*Classification*)

Data yang telah melewati tiga tahap sebelumnya diklasifikasikan berdasarkan standar tertentu yang telah ditentukan sebelumnya. Ini membantu dalam mengorganisir data dengan baik dan memastikan data memenuhi kriteria yang sesuai. Dalam konteks penjualan makanan kuaci, ini dapat diilustrasikan dengan mengklasifikasikan masyarakat ke dalam kelompok tertentu yang layak atau tidak layak menerima penawaran khusus kuaci berdasarkan penghasilan mereka.

#### 5. Pengklusteran (*Clustering*)

Langkah berikutnya adalah mengelompokkan data yang telah diklasifikasikan berdasarkan kesamaan karakteristik. Proses

pengklusteran ini tidak memiliki tujuan tertentu, hanya berfokus pada kesamaan pola atau karakteristik data. Semakin mirip karakteristik antara dua data, semakin tinggi nilai mereka, dan sebaliknya. Dalam konteks penjualan makanan kuaci, ini dapat berarti mengelompokkan bahan bangunan yang diprediksi akan memiliki kenaikan penjualan yang cenderung naik ke dalam kelompok tertentu.

#### 6. Asosiasi (*Association*)

Tahap akhir adalah mengasosiasikan informasi yang dapat melengkapi atau mendukung data yang telah ada, sehingga menghasilkan keputusan yang lebih lengkap. Dalam konteks penjualan makanan kuaci, ini bisa berarti mendapatkan umpan balik dari pembeli yang telah berhasil atau tidak berhasil terjual kuaci, dan menggunakan informasi tersebut untuk meningkatkan penjualan di masa depan.

#### **2.1.1.4 Fungsi *Data Mining***

Menurut penelitian (Setiyani et al., 2020) *data mining* memiliki beberapa fungsi utama, termasuk:

##### 1. Penggalan Informasi

Fungsi *data mining* yang pertama adalah untuk menggali informasi yang tersembunyi atau tidak terlihat secara langsung dari dalam data. Hal ini dilakukan dengan menggunakan teknik analisis statistik, matematika, dan kecerdasan buatan untuk mengidentifikasi pola atau hubungan yang mungkin tidak terlihat dengan mata telanjang.

##### 2. Pengenalan Pola

*Data mining* membantu mengenali pola-pola yang mungkin sulit atau bahkan tidak mungkin ditemukan secara manual. Ini dapat membantu dalam membuat prediksi dan mengidentifikasi tren yang mungkin berguna untuk pengambilan keputusan di masa depan.

### 3. Penyaringan Informasi

*Data mining* membantu menyaring informasi yang relevan dari sejumlah besar data. Dengan cara ini, proses pengambilan keputusan dapat menjadi lebih efisien karena hanya informasi yang penting atau relevan yang dipertimbangkan.

### 4. Klasifikasi dan Prediksi

*Data mining* memungkinkan klasifikasi data ke dalam kategori-kategori yang berbeda dan juga dapat digunakan untuk membuat prediksi berdasarkan pola-pola yang ditemukan dalam data historis.

### 5. Optimasi Proses Bisnis

Dengan menganalisis data operasional, *data mining* dapat membantu meningkatkan efisiensi dan produktivitas proses bisnis dengan mengidentifikasi area yang memerlukan perbaikan atau optimisasi.

### 6. Analisis Kepuasan Pelanggan

Melalui analisis data pelanggan, *data mining* dapat membantu memahami preferensi pelanggan, perilaku belanja, dan kebutuhan pelanggan untuk meningkatkan kepuasan pelanggan.

### **2.1.2 Teori Khusus**

Penelitian ini juga melibatkan pembahasan tentang beberapa prinsip dasar atau landasan argumentasi yang secara khusus terkait dengan variabel-variabel yang menjadi fokus penelitian. Penelitian ini secara khusus mencakup klarifikasi definisi-definisi dari variabel-variabel tersebut.

#### **2.1.2.1 Penjualan**

Penjualan melibatkan pengetahuan dan keterampilan dalam memengaruhi individu, yang dilakukan oleh seorang penjual dengan tujuan mengajak orang lain untuk merespon positif terhadap penawaran produk atau layanan yang diberikan (Pradiani Theresia, 2017). Penjualan merujuk pada tindakan pertukaran barang atau properti dengan imbalan uang. Dalam konteks usaha, ini mencakup proses dimana suatu entitas menyediakan barang atau jasa dan menerima pembayaran uang sebagai hasilnya. Entitas yang terlibat dalam perdagangan barang dan jasa juga bisa dianggap sebagai pihak yang beroperasi dalam bisnis transaksi jual beli. Hasil dari transaksi ini dapat digunakan atau didistribusikan kembali oleh entitas yang telah dijelaskan sebelumnya.

#### **2.1.2.2 Pohon Keputusan**

Pohon keputusan adalah teknik terkenal untuk klasifikasi dan prediksi yang memiliki kekuatan luar biasa. Metode ini mentransformasikan dataset besar menjadi struktur pohon, menggambarkan aturan-aturan dengan cara yang mudah dipahami menggunakan bahasa alami (Muzakir & Wulandari, 2016). Pohon keputusan adalah model prediktif yang menggambarkan keputusan dan konsekuensi berdasarkan serangkaian aturan. Model ini memecah dataset menjadi



bagian-bagian kecil dengan pertanyaan-pertanyaan biner, membentuk struktur pohon. Setiap cabang pohon mewakili keputusan berdasarkan fitur-fitur data, dan daun pohon memberikan prediksi atau nilai. Pohon keputusan dapat digunakan untuk klasifikasi dan prediksi dalam berbagai bidang, dan kelebihanannya meliputi kemudahan interpretasi dan implementasi.

### 2.1.2.3 Algoritma C4.5

Algoritma C4.5, yang dikembangkan oleh Ross Quinlan, adalah algoritma pembangunan pohon keputusan untuk tugas klasifikasi dalam *data mining* dan machine learning. Algoritma C4.5 memanfaatkan konsep *gain* informasi untuk memilih atribut terbaik pada setiap langkah. *Gain* informasi mengukur sejauh mana suatu atribut meningkatkan homogenitas kelas dalam subset data. Algoritma ini dapat menangani data kategorikal dan numerik serta dapat digunakan untuk tugas klasifikasi multi-kelas. Algoritma C4.5 adalah suatu algoritma yang digunakan untuk konstruksi pohon keputusan dalam pengambilan keputusan. Metode ini memilih atribut akar berdasarkan nilai *gain* tertinggi dari atribut yang ada. Hasil model yang dihasilkan dapat berupa aturan IF-THEN, struktur pohon keputusan, rumus matematika, atau model jaringan saraf tiruan. Dengan cara ini, metode C4.5 memungkinkan pengenalan pola dan ekstraksi keputusan dari data yang ada melalui representasi berupa pohon keputusan (Yusuf Maulana et al., 2022). Dalam langkah awal *decision tree*, Proses pemilihan data atau atribut data dilakukan dengan cara mencari nilai *gain* tertinggi dari total nilai setiap data atau atribut. Rumus untuk menentukan nilai *gain* tertinggi adalah:

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

**Rumus 2.1** Perhitungan *Gain*

Penjelasan simbolnya adalah sebagai berikut:

$S$  merupakan himpunan

$A$  merupakan atribut

$n$  merupakan jumlah partisi bagian  $A$

$|S_i|$  merupakan jumlah kasus partisi yang ke- $i$

$|S|$  merupakan jumlah dari kasus dalam  $S$

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

**Rumus 2.2** Perhitungan Entropy

Penjelasan simbolnya adalah sebagai berikut:

$S$  merupakan himpunan

$A$  merupakan fitur

$n$  merupakan jumlah partisi bagian  $S$

$p_i$  merupakan proporsi dari  $S_i$  kepada  $S$

**2.1.2.4 Waikato Environment for Knowledge Analysis (WEKA)**

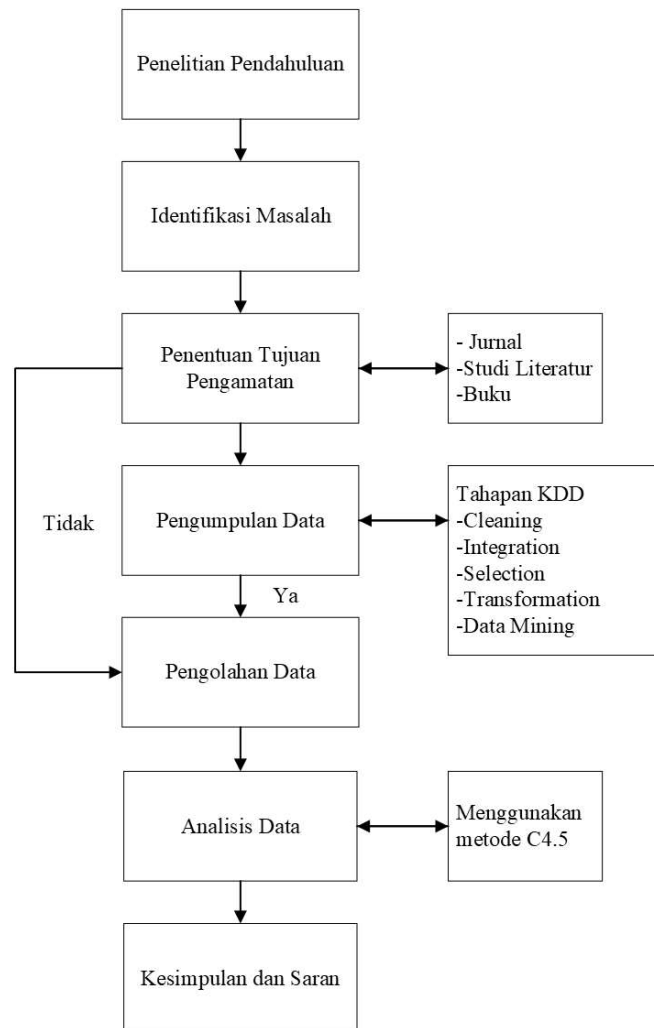
*Waikato Environment for Knowledge Analysis*, atau yang lebih dikenal sebagai WEKA, adalah kumpulan alat praktis untuk pembelajaran mesin. WEKA pertama kali dikembangkan di Universitas Waikato di Selandia Baru dan masih terus diperbarui hingga sekarang, sehingga tetap menjadi alat yang sangat berharga dalam penelitian. WEKA merupakan perangkat lunak sumber terbuka yang dapat

diakses dan diinstal oleh siapa saja di perangkat keras mereka, seperti komputer, laptop, atau tablet (Azwanti, 2018).

WEKA berfungsi sebagai alat bantu pemecahan masalah untuk penelitian dalam bidang *data mining* dan pengklasifikasian masalah yang melibatkan pembelajaran mesin. Penulis menggunakan WEKA versi 3.8.5 menyediakan beragam tahapan analisis *data mining* dan juga memiliki kemampuan untuk memvisualisasikan hasilnya. Hasil uji yang dihasilkan oleh WEKA memiliki fleksibilitas dan efektivitas yang signifikan, terutama ketika diterapkan pada pembuatan *decision tree*. Pohon keputusan yang dihasilkan oleh WEKA cenderung memiliki tingkat kemiripan hasil yang tinggi dibandingkan dengan pengujian manual.

## **2.2 Kerangka Pemikiran**

Penelitian ini mengembangkan suatu struktur konseptual yang membentuk landasan berpikir, menghasilkan suatu rangkaian langkah logis yang diterapkan oleh peneliti untuk membuktikan dan akhirnya mencapai hasil yang menjadi dasar pengambilan keputusan. Proses ini dapat disusun dan dianalisis lebih lanjut melalui representasi visual dalam bentuk diagram, seperti yang terlihat pada gambar di bawah ini:



**Gambar 2.2** Kerangka Berpikir

**Sumber:** (Peneliti, 2023)

### 2.3 Hipotesis Penelitian

Hipotesis dapat diartikan sebagai perkiraan sementara yang mengantisipasi kemungkinan hubungan antara satu variable dengan variable lainnya melalui serangkaian uji, yang kemudian akan mengarah kepada temuan yang valid atau kebenaran dari permasalahan yang sedang diteliti. Dalam konteks ini, peneliti telah merumuskan hipotesis awal penelitian yang mencakup:

1. Duga bahwa penerapan teknik *data mining* klasifikasi menggunakan algoritma C4.5 memiliki kemampuan untuk menganalisis pola penjualan produk makanan kuaci di PT Prima Niaga Indomas.
2. Duga bahwa hasil analisis dari penelitian ini memiliki potensi untuk memberikan prediksi yang akurat terkait dengan penjualan produk makanan kuaci di PT Prima Niaga indomas.

#### **2.4 Penelitian Terdahulu**

Bagian ini berisi gambaran mengenai penelitian-penelitian sebelumnya yang telah dilakukan oleh berbagai peneliti dengan fokus pada analisis menggunakan algoritma C4.5. Hasil-hasil dari penelitian-penelitian ini mencakup topik yang relevan atau berhubungan dengan pemahaman algoritma C4.5, antara lain:

1. Dalam penelitian berjudul “Analisa algoritma c4.5 untuk memprediksi penjualan motor pada PT. Capella Dinamik Nusantara Cabang Muka” yang dilakukan oleh (Azwanti, 2018) ditemukan bahwa penerapan algoritma C4.5 berhasil terbukti memberikan dukungan signifikan terhadap sektor penjualan perusahaan. Algoritma ini efektif dalam menganalisis dan menghasilkan prediksi-prediksi yang berharga terkait aktivitas penjualan motor di PT. Capella Dinamik Nusantara.
2. Dalam penelitian berjudul “Penerapan *data mining* untuk prediksi penjualan *wallpaper* menggunakan algoritma C4.5” yang dilakukan oleh (Juna Eska, 2016) ditemukan bahwa penerapan *data mining*, khususnya melalui algoritma C4.5, dapat memberikan manfaat yang besar dalam meningkatkan kepuasan pembelian *wallpaper*. Dengan menggunakan

metode ini, proses pengambilan keputusan dalam pembelian wallpaper dapat ditingkatkan secara efektif.

3. Dalam penelitian berjudul “Penerapan k-means *clustering* pada data penerimaan mahasiswa baru” yang dilakukan oleh (Nasari & Darma, 2015) ditemukan bahwa penggunaan metode *data mining*, khususnya teknik kluster K-Means, dapat digunakan untuk mengidentifikasi kecenderungan jurusan yang mungkin dipilih oleh calon mahasiswa setelah menyelesaikan pendidikan di sekolah menengah.
4. Dalam penelitian berjudul “Penerapan metode clustering dengan algoritma k-means pada pengelompokkan data calon mahasiswa baru di Universitas Muhammadiyah Yogyakarta” yang dilakukan oleh (Asroni et al., 2018) ditemukan bahwa penerapan metode *data mining*, khususnya dengan menggunakan algoritma kluster K-Means, dapat digunakan untuk mengidentifikasi preferensi calon mahasiswa baru terhadap program studi Pendidikan Dokter dan Ilmu Hubungan Internasional di Universitas Muhammadiyah Yogyakarta. Hasil penelitian ini memberikan dasar bagi pihak universitas dalam merancang strategi promosi untuk menarik minat calon mahasiswa baru ke jurusan-jurusan tersebut.
5. Dalam penelitian berjudul “Implementasi *data mining* dalam prediksi kepuasan belajar saat pandemic covid menggunakan algoritma c4.5” yang dilakukan oleh (Siahaan & Fauzi, 2023) ditemukan bahwa pohon keputusan memiliki potensi sebagai panduan bagi lembaga pendidikan untuk meningkatkan kepuasan belajar siswa selama proses pembelajaran.

Berdasarkan hasil uji sistem, dapat disetujui bahwa aplikasi ini berhasil meningkatkan tingkat kepuasan siswa dalam pembelajaran online selama periode pandemi COVID-19.

6. Dalam penelitian berjudul “analisis prediksi kelulusan mahasiswa tepat waktu menggunakan metode *data mining* naïve bayes : systematic review” yang dilakukan oleh (Setiyani et al., 2020) ditemukan bahwa melalui identifikasi literatur yang dilibatkan, metode *data mining* naïve bayes dapat digunakan untuk memprediksi kelulusan mahasiswa tepat waktu dengan mempertimbangkan berbagai atribut dari basis data perguruan tinggi. Meskipun terdapat variasi dalam jumlah atribut dan aplikasi *data mining* yang digunakan dalam ketiga literatur, namun tingkat akurasi yang dihasilkan konsisten di atas 90%. Salah satu atribut yang secara konsisten mempengaruhi prediksi adalah Indeks Prestasi Kumulatif (IPK).
7. Dalam penelitian berjudul “*Data mining* : klasifikasi menggunakan algoritma c4.5” yang dilakukan oleh (Mardi Yuli, 2017) ditemukan bahwa seluruh proses *data mining* mengarah pada pembentukan pohon keputusan yang memberikan informasi yang diperlukan. Tahap-tahap utama dalam penelitian ini melibatkan sumber data, yang merupakan database yang berisi informasi yang dapat diekstrak dan dimanfaatkan untuk kepentingan bisnis dan penelitian.