

BAB II

TINJAUAN PUSTAKA

2.1. Teori

Pada bagian ini merupakan pembahasan mengenai teori yang digunakan, teori merupakan suatu fakta dan mekanisme yang sudah jelas dan bersifat ilmiah yang memiliki peran untuk mendapatkan penjelasan dan mengelompokkan pengalaman.

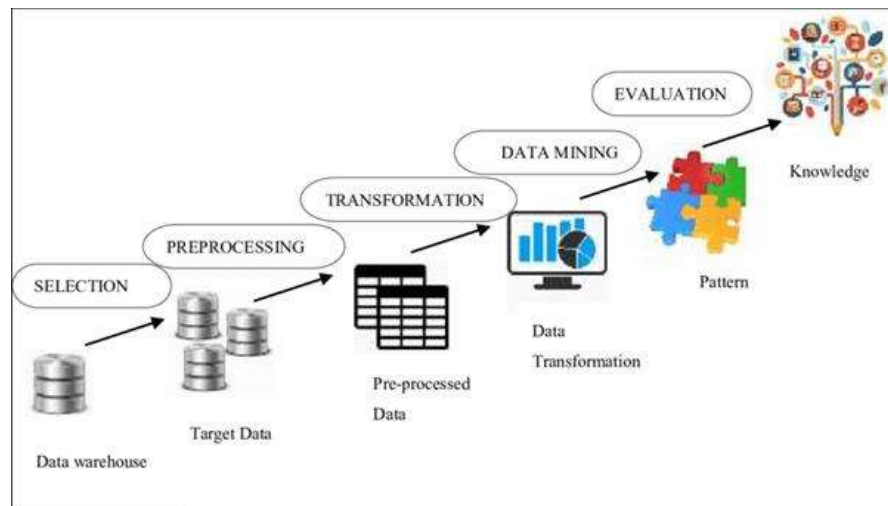
2.1.1. Teori Umum

Teori umum adalah suatu landasan dalam sebuah penelitian yang menjadi teori pokok.

2.1.1.1. *Knowledge Discovery in Database*

Knowledge discovery in database ini sering dikaitkan dengan data mining karena memiliki arti yang sama yaitu pengumpulan data dalam set data yang berukuran besar kemudian diolah dengan sebuah proses menjadi sebuah informasi yang berharga (Yusuf Maulana et al., 2022). Data yang berjumlah besar sangat berpotensi datanya menghilang dan tidak konsisten kemudian hal tersebut akan sangat berpengaruh pada proses data mining (Alasadi, 2017).

Untuk lebih jelasnya dapat dilihat dari gambar berikut.



Gambar 2.1 *Knowledge Discovery in Database*

Sumber : (Researchgate, 2023)

Berikut merupakan untuk langkah – langkah dari *knowledge discovery in database* (Andi fahlevi, 2021).

1. *Data Cleansing*, Proses membersihkan data yang kosong dan dianggap tidak penting sehingga tersisa data yang bisa digunakan saja.
2. *Data Integration*, Menggabungkan data yang memiliki nilai sama menjadi satu.
3. *Selection*, Menyeleksi data sehingga data yang digunakan hanya yang relevan terhadap kasus penelitian.
4. *Data Transformation*, Mentransformasikan bentuk data menjadi bagian-bagian yang sesuai untuk diolah.
5. *Data Mining*, Data diolah menjadi sebuah pola – pola sesuai aturan yang sudah ada dan digunakan.

6. *Pattern Evolution*, Pola yang dihasilkan diidentifikasi berdasarkan ukuran yang tersedia.
7. *Knowledge Presentation*, Data yang sudah selesai diolah dan mengeluarkan hasil dibentuk menjadi sebuah tampilan yang lebih menarik dan dapat dimengerti orang awam.

2.1.1.2. Data Mining

Data mining merupakan proses pengumpulan data kemudian dianalisis beberapa bagian dari kumpulan data besar yang sudah ada setelah itu dikembangkan menjadi sebuah data baru yang bermanfaat bagi penggunanya (Clifton, 2010). Teknik ini sangat menguntungkan dipakai untuk menyelidiki sebuah informasi dan mengembangkannya menjadi data yang penting dan bermanfaat. Informasi yang ditambang dari *database* besar menurut banyak peneliti dianggap sebagai topik penelitian utama dalam sistem database yang penting dan memiliki peluang pendapatan yang besar (Ming-Syan Chen, 2022). Kemajuan teknologi telah mengubah cara menangani data. Sekarang digunakan untuk berbagai kumpulan data untuk mempelajari pola tersembunyi dan membuat prediksi atau deskripsi yang tepat. Penambangan data adalah kumpulan metode yang digunakan untuk mengekstrak pengetahuan tersembunyi seperti pola, hubungan, atau aturan dari kumpulan data yang besar (Mienye et al., 2019). *Data mining* juga menjadi salah satu metode penyelidikan yang berkembang sangat cepat dalam dunia teknologi informasi dikarenakan proses pengumpulan informasi yang diolah dengan baik untuk menunjang keputusan yang krusial di masa yang akan datang membuatnya

menjadi metode yang sangat dipertimbangkan untuk digunakan.

2.1.1.3. Klasifikasi

Klasifikasi data merupakan langkah untuk menemukan bagian sama dalam basis data dan memisahkannya ke dalam kelas yang berbeda menurut model yang ditetapkan (Azwanti, 2018). Klasifikasi ini sendiri erat kaitannya dengan algoritma C4.5 dikarenakan mampu menghasilkan *decision tree* yang diimplementasikan dan diambil sebuah keputusan di masa yang akan datang (Angraini et al., 2018).

2.1.1.4. Algoritma C4.5

Algoritma C4.5 yang dibuat oleh J.Rose Quinlan merupakan perkembangan dari perhitungan ID3 yang merupakan ekstensi nya dengan prinsip decision tree yang hampir sama. Hal ini merupakan proses klasifikasi data dengan sistem kerja penyusunan langkah yang sesuai untuk mengatasi sebuah masalah (Nafiyah, 2020). Dengan cara yang sama seperti *Dichotomizer 3* (ID3) berulang, itu menggunakan konsep *entropi* informasi dari serangkaian pelatihan data untuk membangun pohon keputusan. Data pelatihan terdiri dari sekumpulan sampel yang telah diklasifikasikan sebelumnya. C4.5 di setiap simpul pohon (Lo et al., 2019).

Berikut beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5:

1. Menyiapkan data *training*. Data ini berasal dari data sebelumnya yang telah dikelompokkan kedalam kelas tertentu.
2. Selanjutnya, temukan akar pohon atau *node*. Anda dapat melakukannya

dengan menghitung nilai *gain* dari setiap atribut, dengan akar pertama menghasilkan nilai *gain* yang paling besar. Sebelum mengetahui nilai *gain*, hitung nilai *entropy* terlebih dahulu.

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

Rumus 2.1 Rumus *entropy*

- Setelah itu menentukan nilai *gain* menggunakan rumus:

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

Rumus 2.2 Rumus *Gain*

Keterangan :

S = himpunan kasus

A = atribut

n = jumlah partisi atribut A

|S_i| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

- Ulangilah langkah mencari *entropi* hingga seluruh rekor data terbagi sempurna.

2.1.1.5. Decision Tree

Decision tree adalah alat pendukung keputusan yang direpresentasikan dalam bentuk struktur pohon yang memiliki *node*, *branch* dan *leaf*. *Decision tree* adalah alat yang mendukung dengan struktur seperti pohon yang mendefinisikan berbagai probabilitas atau hasil dari sebuah pernyataan. Sebuah pohon keputusan merupakan rangka sederhana yang dihasilkan melalui salah satu metode yaitu

algoritma C4.5 pada setiap *node non-terminal* adalah tes dan dianggap menjadi item data (Anggraini et al., 2018). *Decision tree* membantu menyajikan algoritma dengan pernyataan kontrol bersyarat. Mereka terdiri dari subbagian yang menunjukkan proses pengambilan keputusan yang dapat menghasilkan hasil yang menguntungkan. Struktur *flowchart* memiliki *node* internal yang mewakili tes atau fitur pada setiap tahap proses. Setiap cabang menunjukkan hasil atribut, dan jalur dari daun ke akar menunjukkan aturan klasifikasi. Mereka meningkatkan akurasi, stabilitas, dan kemudahan interpretasi model prediktif. Karena alat ini dapat menangani masalah penyesuaian data seperti regresi dan klasifikasi, itu juga berguna untuk menyesuaikan hubungan non-linier (Kaufmann, 2022).

2.1.1.6. *Rapid Miner*

Rapidminer merupakan salah satu perangkat lunak yang dikembangkan untuk membantu proses perhitungan dalam data mining. Aplikasi ini lebih banyak digunakan dalam *machine learning*, analisis prediktif dan pembelajaran mendalam lainnya (Nofitri & Irawati, 2019). *Rapidminer* merupakan bagian dari perangkat lunak untuk mengolah data dan membuat bentuk dari set data yang besar menggunakan salah satu prinsip dalam data mining dan penggabungan metode lainnya. Aplikasi ini sangat membantu dalam perhitungan data set yang besar kemudian menghasilkan hasil dalam waktu yang lebih singkat (Wijaya et al., 2023).

2.1.2. Teori Khusus

Teori khusus adalah teori yang menjelaskan fakta dalam hubungannya satu sama lain.

2.1.2.1. E-Commerce

E-commerce merupakan proses transaksi pembelian barang yang terjadi secara online di sebuah platform yang menyediakan barang tersebut (Azizah & Aswad, 2022). Hal ini sangat populer terutama di kalangan kaum muda milenial. Akibatnya sampai hari ini banyak sekali anak muda yang lebih memilih belanja lewat salah satu platform e-commerce daripada belanja di toko, supermarket atau sejenisnya karena dinilai lebih mudah aksesnya, tidak perlu repot bepergian, dan banyak promonya. Fenomena inilah yang mengubah perekonomian di Indonesia. Pembayaran yang dilakukan dengan berbagai metode tanpa tunai (akses *online*) juga membuat perputaran uang semakin cepat dari pembeli ke penjual di berbagai daerah bisa dijangkau dengan mudah (Harahap & Amanah, 2018).

2.1.2.2. Penjualan

Penjualan merupakan proses yang melibatkan penawaran dalam bentuk produk ataupun jasa dari penjual kepada pembeli dengan nilai tukar tertentu (Elisa, 2022). Penjualan juga dapat dipahami sebagai kegiatan yang bertujuan memenuhi kebutuhan konsumen dan mendapatkan *profit* dari selisih *margin* modal (Ade Izyuddin & Setyawan Wibisono, 2020), dalam kasus marketplace ini adalah penjualan sebuah produk yang dipasarkan online kemudian dapat dilihat oleh calon pelanggan dengan cara mencari melalui kata kunci yang sesuai. Transaksi yang

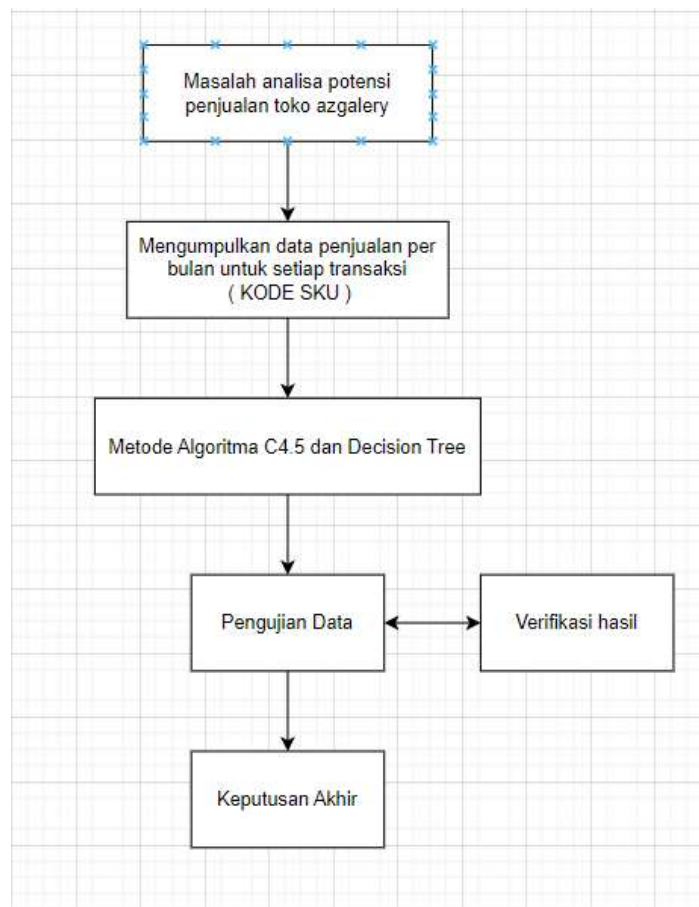
terjadi Ketika pelanggan membeli barang dengan berbagai macam jenis pembayaran, hal itulah yang dikatakan penjualan. Penjualan di marketplace dapat dipengaruhi oleh berbagai faktor yang dapat memicu terjadinya penjualan. Faktor dasar yang dapat memicu adanya penjualan tentunya tampilan produk yang menarik disertai video yang dapat meyakinkan pelanggan. Dikarenakan penjualannya terjadi dalam sistem online, penjual harus menumbuhkan rasa percaya kepada calon pembeli agar tertarik membeli produk tersebut. Salah satu faktor dasar lainnya yang paling penting tentunya kata kunci yang membuat produk dari penjual bisa ditemukan setelah pembeli mencari kata kunci yang relevan dari produk yang diminatinya.

2.1.2.3. Produk

Produk merupakan barang atau jasa yang dijual kepada pelanggan. Produk bisa dikatakan sebagai nilai tukar yang dimiliki oleh penjual yang ditawarkan kepada pembeli. Produk di *marketplace* banyak jenis dan kategorinya dan cakupannya luas dari berbagai daerah (Purnama & Putri, 2021). Produk yang paling banyak menarik minat di *marketplace* adalah *fashion*, dekorasi kemudian peralatan rumah tangga karena hal tersebut dinilai praktis dan lebih memiliki potensi penjualan yang baik daripada produk berkategori elektronik, otomotif, kesehatan dikarenakan cenderung ditawarkan dengan harga yang mahal dan butuh tingkat kepercayaan dari calon pembeli yang tinggi (Maitri Priandewi, 2021). Sehingga banyak pembeli kemungkinan membeli kategori produk tersebut *offline* saja karena bisa melihat produknya langsung. Tidak semua produk juga bisa dijual di

marketplace karena umumnya memiliki aturan yang ketat mengenai produk yang dilarang salah satunya adalah produk *import* yang memiliki indikasi palsu atau membahayakan seperti senjata tajam.

2.2. Kerangka Pemikiran



Gambar 2.2 Kerangka Pemikiran

Sumber : (Penulis, 2023)

2.3. Hipotesis Penelitian

Hipotesis merupakan dugaan sementara kepada suatu kejadian yang kebenarannya masih perlu diperhitungkan kembali melalui penelitian yang akan dilakukan. Berdasarkan masalah di atas maka penulis membuat hipotesis sebagai berikut :

1. Diduga metode data mining dengan perhitungan algoritma C4.5 mampu dan efektif dalam menganalisa potensi pembeli.
2. Diduga hasil penelitian ini dapat mengetahui potensi penjualan di toko mitra azgalery

2.4. Penelitian Terdahulu

Dibawah ini merupakan penelitian terdahulu yang membantu penelitian yang saya lakukan saat ini yang memiliki keterikatan dengan topik yang dibahas:

1. Jurnal berjudul “Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT. Cappela Dinamik Nusantara Cabang Muka Kuning“ (Azwanti, 2018). Penelitian ini berisi tentang penggunaan metode algoritma C4.5 untuk memprediksi penjualan motor di sebuah perusahaan.
2. Jurnal berjudul “Analisa Prediksi Penjualan Produk Dengan Menggunakan Metode C4.5 (Studi Kasus: PT. Kawan Lama Ace Hardware)” (Lubis, 2019). Membahas mengenai prediksi penjualan 3500 jenis produk dengan algoritma C4.5 pada sebuah perusahaan.
3. Jurnal berjudul “Analisis Data Mining Penjualan Ban Menggunakan Algoritma C4.5”(Anggraini et al., 2018). Meneliti penjualan berdasarkan jenis ban dengan kriteria klasifikasi banyak dan sedikit kemudian diuji

tingkat akurasi datanya.

4. Jurnal berjudul “Aplikasi Prediksi Penjualan AC Menggunakan Decision Tree Dengan Algoritma C4.5” (Ade Izyuddin, 2020). Penelitian yang dilakukan untuk menyelesaikan masalah stok yang sering kosong saat situasi puncak penjualan yang sedang tinggi.
5. Jurnal berjudul “Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Memprediksi Penjualan Alat – Alat Telekomunikasi Dan IT (Information Technology) Pada PT Julian Adiputra Utama” (Wijaya et al., 2023) . Membahas mengenai pemanfaatan metode algoritma C4.5 untuk menentukan jumlah stok yang tepat untuk sebuah produk agar tidak terjadi penumpukan stok di gudang.
6. Jurnal berjudul “Implementasi Algoritma C4.5 Untuk Klasifikasi Penjualan Barang di Swalayan Dutalia” (Kristianus Lalo et al., 2021). Penelitian yang menemukan pola penjualan barang untuk dapat memprediksi kondisi pasar sehingga mendapatkan angka penjualan prediksi kedepannya yang bisa dipakai sebagai acuan.
7. Jurnal berjudul “Penerapan Algoritma C4.5 pada Analisis Data Penjualan Obat Herbal Untuk Prediksi Penjualan Di Toko Online XXX” (Adat et al., 2022). Membahas mengenai prediksi penjualan obat herbal.
8. Jurnal berjudul “Penerapan Algoritma C4.5 Untuk Memprediksi Penjualan Lele Pada Kolam Pancing Galatama” (Diansyah & Exprada, 2022). Penelitian ini membahas mengenai analisa jenis ikan lele yang penjualannya bagus untuk mempercepat pengambilan keputusan.

9. Jurnal berjudul “Penerapan Data Mining Menggunakan Algoritma C4.5 Terhadap Pengaruh Penjualan Kopi Pada PT.JPW Indonesia” (Ifongki, 2020). Membahas mengenai solusi untuk penjualan yang kurang stabil akibat kurangnya pengolahan data yang berpotensi untuk pemanfaatan pengambilan keputusan.
10. Jurnal berjudul “Prediksi Potensi Penjualan Makanan Beku berdasarkan Ulasan Pengguna Shopee menggunakan Metode *Decision Tree* Algoritma C4.5 dan *Random Forest* (Studi Kasus Dapur Lilis)” (Ferdyandi et al., 2022). Membahas mengenai prediksi untuk mengetahui potensi produk makanan beku yang sesuai dengan minat pelanggan berdasarkan ulasan.