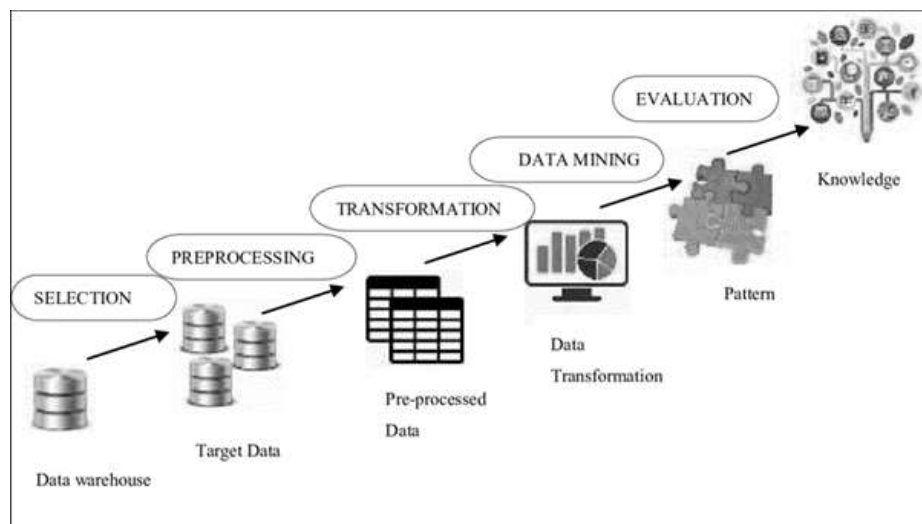


BAB II

TINJAUAN PUSTAKA

2.1 *Knowledge Discovery in Database (KDD)*

Menurut Rajput (2023), *Knowledge Discovery in Database (KDD)* adalah proses yang melibatkan ekstraksi informasi yang berguna, sebelumnya tidak diketahui, dan informasi yang memiliki potensi berharga dari kumpulan data. Muslim et al. (2019) memaparkan proses KDD dalam *data mining* seperti pada gambar 2.1 berikut:



Gambar 2.1 Proses KDD
Sumber: (Peneliti, 2023)

1. *Selection*

Sebelum dilakukan pengujian, akan dilakukan pengolahan berupa seleksi data agar hasil pengujian lebih akurat.

2. *Pre-Processing*

Proses dimana data dibersihkan dan diubah agar siap untuk dianalisis.

Proses ini termasuk normalisasi data, penanganan data yang hilang, dan integrasi data.

3. *Transformation*

Transformasi merupakan tahap mengubah data dan menggabungkan data ke dalam format tertentu, seperti matriks atau grafik.

4. *Data Mining*

Proses dimana terjadinya penerapan teknik dan algoritme *data mining* ke data untuk mengekstrak informasi yang bermanfaat. Proses ini diantaranya pengelompokan, klasifikasi, *association rule mining*, dan deteksi anomali.

5. *Evaluation*

Mengevaluasi hasil untuk memastikan bahwa hasil data yang diekstraksi berguna, akurat, dan bermakna.

6. *Knowledge Presentation*

Proses penyajian tahap akhir dari proses *data mining* yaitu merumuskan aturan asosiasi dari hasil analisis yang diperoleh.

2.2 Data Mining

Data mining merupakan proses menemukan korelasi, pola, dan tren dengan menyaring sejumlah besar data yang disimpan dalam repositori, menggunakan pola teknologi pengenalan serta teknik statistik dan matematika (Larose & Larose, 2014). Vlandari (2017) menjelaskan bahwa *data mining* merupakan rangkaian proses untuk menambah dan mencari informasi yang belum diketahui secara

manual dari sebuah *database*. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola-pola penting atau menarik dari data yang terdapat dalam *database*. Menurut Mustika et al. (2021), *data mining* merupakan sebuah teknologi yang dapat memproses data dalam volume besar yang digunakan oleh perusahaan untuk mengubah data mentah menjadi informasi yang berguna untuk membuat suatu keputusan bisnis di masa yang akan datang. *Data mining* mempunyai tujuan sebagai *explanatory* yaitu untuk menjelaskan beberapa kondisi terkait dengan suatu penelitian dan sebagai *confirmatory* untuk mempertegas hipotesis.

2.3 Metode Data Mining

Menurut Muslim et al. (2019), ada 5 (lima) metode yang digolongkan dalam data mining diantaranya:

1. *Estimation* (Estimasi)

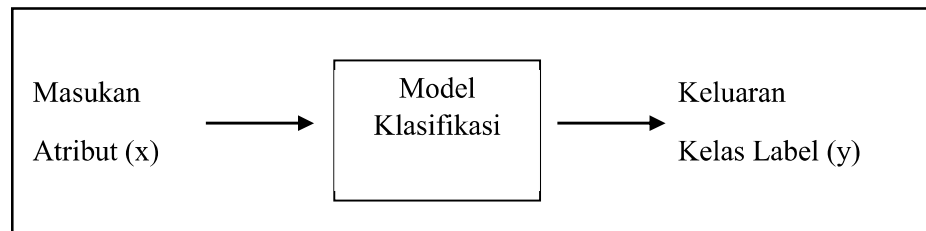
Estimasi merupakan metode yang mampu memperkirakan nilai variabel target numerik menggunakan satu set variabel prediktor numerik dan kategorikal.

2. *Forecasting* (Prediksi/Peramalan)

Metode ini hampir sama dengan estimasi, perbedaannya terletak pada hasil prediksi yang terletak di masa yang akan datang.

3. *Classification* (Pengelompokan)

Klasifikasi merupakan metode yang digunakan untuk mendeskripsikan konsep atau kelas data. Model klasifikasi dapat dilihat pada Gambar. 2.2 berikut:



Gambar 2.2 Blok Diagram Model Klasifikasi
Sumber: (Peneliti, 2023)

4. *Clustering* (Klastering)

Klastering merupakan pengelompokkan data secara otomatis atau sering disebut sebagai metode *unsupervised learning*.

5. *Association rule* (Aturan Asosiasi)

Teknik ini berfokus untuk menemukan aturan kesamaan dalam suatu kejadian, seperti pembelian barang dagangan pada pusat perbelanjaan. Metode asosiasi yang sering digunakan yaitu *Chi Square*, *FP-Growth*, *A Priori*, dll. Terdapat dua metode dasar *association rule*, diantaranya:

a. Analisa pola frekuensi tinggi

Tahapan ini merupakan tahap menentukan seberapa sering *itemset* muncul di *dataset*. Di tahap ini perlu mencari nilai *support* untuk menemukan hubungan antara produk dengan keseluruhan kumpulan data. Berikut persamaan untuk mencari nilai *support* dari sebuah *itemset*:

$$SUPPORT = \frac{\sum \text{transaksi mengandung } A}{\sum \text{transaksi}} \times 100\%$$

Rumus 2.1 Persamaan Minimum *Support*

Keterangan:

- *Support* merupakan persentase nilai kombinasi *itemset* dalam total transaksi.
- Transaksi mengandung A merupakan jumlah kemunculan item A dalam transaksi.

b. Pembentukan aturan asosiasi

Setelah menentukan *support* dalam *itemset*, tahap selanjutnya yaitu menentukan *confidence* untuk melihat hubungan antara satu atau lebih item dan item lainnya. Berikut persamaan rasio transaksi yang mengandung A dan B terhadap jumlah total nilai A yang ada, di mana A adalah anteseden dan B adalah konsekuen:

$$Confidence = P(A | B) = \frac{\sum \text{transaksi A \& B}}{\sum \text{transaksi A}} \times 100\%$$

Rumus 2.2 Persamaan *Confidence*

Keterangan:

- *Confidence* merupakan tingkat kepastian hubungan antar item yang muncul secara bersamaan.
- Transaksi A & B merupakan total transaksi yang mengandung item A & B.

2.4 *Software* Pendukung

Salah satu *software* pendukung dalam pengelolaan *data mining* yaitu *Rapidminer*. *Software* ini tergolong *open source* untuk penelitian *text mining*, *data mining*, dan analisis prediksi. Operator yang terdapat pada *RapidMiner* diantaranya

input, output, visualisasi, dan data preprocessing. Sebagai *software open source*, *RapidMiner* tidak perlu diragukan lagi karena *software* ini sudah terkemuka di dunia (Muslim et al., 2019). Berikut aplikasi yang digunakan dalam pengelolaan data mining seperti pada gambar. 2.3 berikut:



Gambar 2.3 Software Rapidminer
Sumber: (Peneliti, 2023)

2.5 Penelitian Terdahulu

Penelitian ini didukung oleh beberapa penelitian sebelumnya, salah satunya yang dilakukan oleh Panjaitan et al. (2019) dengan judul penelitian *Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns*. Penelitian ini menggunakan algoritma apriori dengan *software* pendukung yaitu Visual Basic 2010. Hasil penelitian yang ditemukan yaitu analisis pola pembelian konsumen menggunakan algoritma apriori pada aplikasi ini, menghasilkan aturan asosiasi dengan minimum *support* 15% dan *confidence* 50% menghasilkan sebanyak 87 *rules*.

Mardianti & Fauzi (2020) melakukan penelitian dengan judul Algoritma Apriori Dalam Menentukan pola Konsumen Terhadap Tata Letak Barang. Penelitian ini menggunakan algoritma apriori dengan *software* pendukung yaitu Rapid miner. Hasil penelitian ini menghasilkan 4 aturan asosiasi dengan *support*

5% dan *confidence* 50% yang dapat menjadi rekomendasi pelaku usaha dalam penyusunan tata letak barang untuk meningkatkan strategi penjualan.

Dewi, Irawan, & Sormin (2022) melakukan penelitian dengan judul Penerapan Algoritma Apriori Terhadap Data Mining Penjualan Di Alfamart Berastagi. Penelitian ini menggunakan metode apriori dengan hasil penelitian diperoleh 75 aturan asosiasi yang terbentuk dengan jumlah minimum *support* 50% dan minimum *confidence* 90% dan aturan asosiasi tertinggi adalah apabila membeli *FRENCH FRIES* maka kemungkinan juga akan membeli *BREAD BREAND ORICAN*.

Agustiani et al. (2020) melakukan penelitian dengan judul Penerapan Data Mining Metode Apriori Dalam Implementasi Penjualan Di Alfamart. Penelitian ini menggunakan bahasa pemrograman *java* dan *Mysql*. Secara keseluruhan dari data sampel penjualan diperoleh 142 *rule* asosiasi yang memenuhi *support* diatas 10% dan *confidence* 50%.

Takdirillah (2020) melakukan penelitian dengan judul Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan. Penelitian ini dilakukan dengan pendekatan *knowledge discovery in database*. Aturan asosiatif yang dibentuk diuji menggunakan *lift ratio*, supaya dapat diketahui aturan asosiasi antar produk mana saja yang paling kuat. Hasil penelitian menunjukkan bahwa algoritma apriori dapat digunakan untuk mengolah data transaksi penjualan menjadi informasi baru berupa keterkaitan antar produk yang didasari dari pengujian dengan tools *Orange*.

Saputra & Sibarani (2020) melakukan penelitian dengan judul Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat. Penerapan Algoritma Apriori dalam aplikasi ini berhasil mencari kombinasi item terbanyak berdasarkan data transaksi dan kemudian membentuk pola asosiasi dari kombinasi item tersebut. Penelitian ini menggunakan metode apriori dengan hasil penelitian dapat mengetahui jenis obat yang sering dibeli oleh konsumen secara bersamaan sehingga dapat mengetahui pola penjualan obat.

Erfina, Melawati, & Arianti (2020) melakukan penelitian dengan judul Penerapan Metode Data Mining Terhadap Data Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus: Toko Fasentro Fancy). Penelitian ini menggunakan metode apriori dengan hasil penelitian ditemukan 4 aturan asosiasi yang memenuhi nilai *minimum support* dan nilai *minimum confidence* yang telah ditentukan. Berdasarkan hasil aturan asosiasi final yang memenuhi nilai *support* dan nilai *confidence* tertinggi yaitu jika Pranaya (P24TVC), Toyobu Fodu maka akan membeli Roberto Cavali.

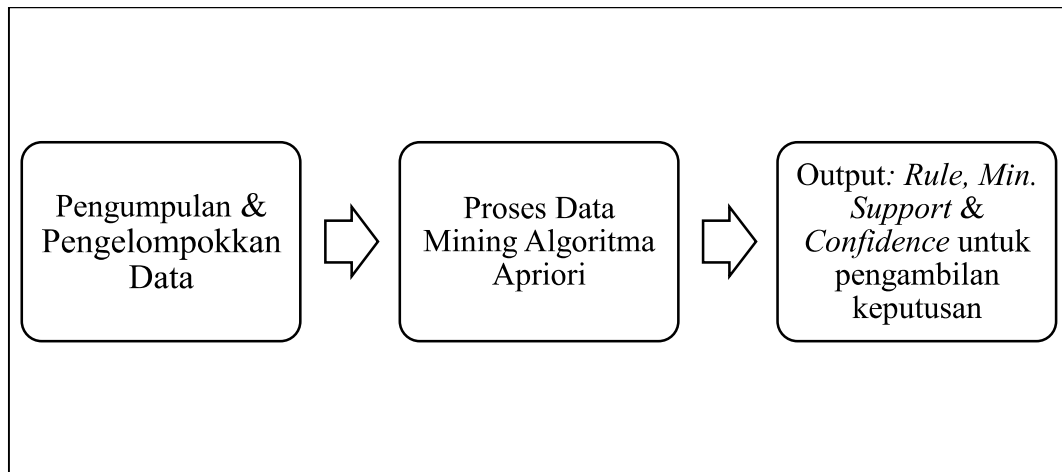
Rizsky & Sadikin (2019) melakukan penelitian dengan judul Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan. Aturan asosiasi yang terbentuk digunakan sebagai acuan untuk rekomendasi produk yang memenuhi nilai *confidence* dan *support minimum*. Hasil dari penelitian ini yaitu algoritma apriori yang diuji cobakan pada data set transaksi penjualan fleksibel untuk digunakan sebagai dasar pengambilan keputusan perusahaan pada area pemasaran.

Susilo (2018) melakukan penelitian dengan judul Penerapan Algoritma Apriori pada Pengolahan Data Transaksi Penjualan di Minimarket Priyo Kota Lubuklinggau. Penerapan algoritma apriori pada proses pengolahan data hasil transaksi penjualan di Minimarket Priyo dapat membentuk beberapa pola kombinasi itemsets dan informasi yang dihasilkan dapat digunakan dalam pembuatan katalog produk yang akan dijual.

Sikumbang (2018) melakukan penelitian dengan judul Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori. Algoritma Apriori membantu mengembangkan strategi penjualan sepatu. Hasil dari penelitian ini ditemukan beberapa merek sepatu yang paling diminati dan hasil ini dapat dijadikan sebagai acuan peletakkan sepatu di toko tersebut.

2.6 Kerangka Pemikiran

Kerangka penelitian berguna untuk menentukan proses dan tahap yang akan dilakukan dalam sebuah penelitian. Berdasarkan penjelasan di atas, penelitian ini dimulai dari tahap pengumpulan dan pengelompokan data berdasarkan item, kemudian data yang sudah diperoleh akan diolah menggunakan Algoritma Apriori. Hasil output berupa *rules* kombinasi, nilai minimum *support* dan *confidence* akan dijadikan acuan pengambilan keputusan oleh manajemen. Kerangka pemikiran pada penelitian ini digambarkan pada gambar 2.4 berikut:



Gambar 2.4 Kerangka Pemikiran

Sumber: (Peneliti, 2023)