

## **BAB II**

### **TINJAUAN PUSTAKA**

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

Metode penemuan pengetahuan dalam Basis Data (KDD) melibatkan pengumpulan dan penggunaan informasi masa lalu dengan tujuan mempelajari data saat ini, dengan tujuan menciptakan model untuk mengidentifikasi lebih banyak pola data serta pola data berskala besar (Sitohang et al. 2022).

Berikut ini adalah langkah-langkah ekstraksi informasi dalam prosedur jaringan (Elisa 2018) :

##### 1. Pembersihan Data

Untuk menghilangkan ketidaksesuaian atau data yang saling bertentangan, pembersihan data dilakukan di awal proses.

##### 2. Integrasi Data

Integrasi data adalah proses penggabungan data dari banyak database ke dalam satu database baru.

##### 3. Pilihan Data

Untuk memilih data yang relevan dari database adalah tujuan dari seleksi data.

##### 4. Proses mengubah data

Proses mengubah data ke dalam format yang dapat digunakan oleh data mining dikenal sebagai transformasi data.

##### 5. Penambangan Data

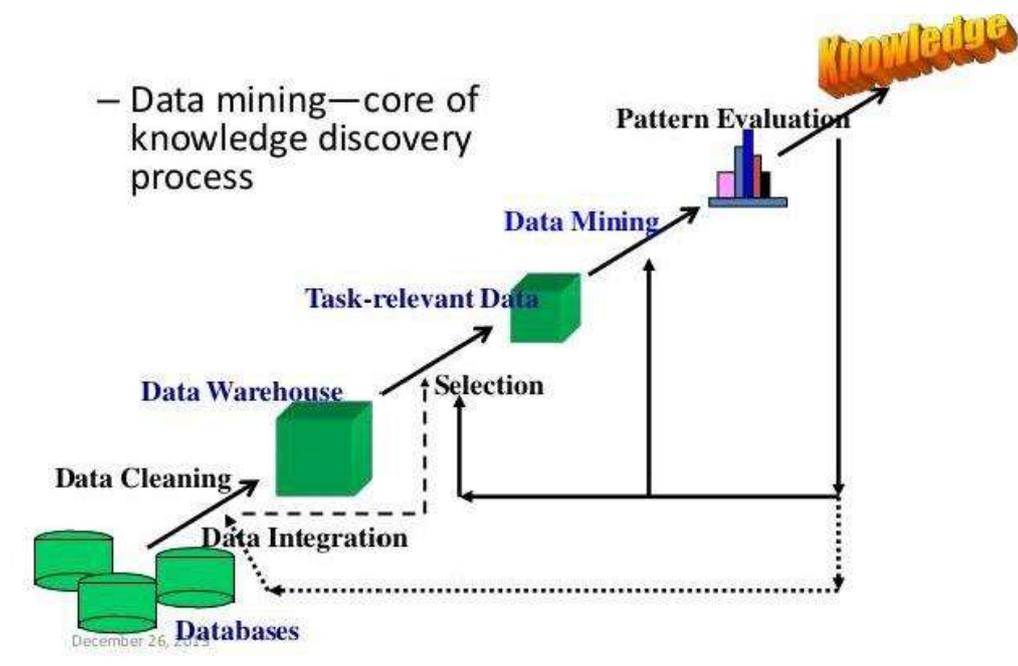
Penambangan data adalah metode untuk menemukan informasi penting yang tersembunyi di dalam data.

## 6. Penilaian Pola

Evaluasi pola untuk menemukan pola yang menarik untuk presentasi berbasis pengetahuan.

## 7. Presentasi Pengetahuan

Tujuan dari presentasi pemahaman adalah untuk memperjelas dan menyampaikan informasi tentang pendekatan yang digunakan orang untuk mempelajari konsep-konsep baru.



**Sumber :** (Yanto and Kesuma 2017)

**Gambar 2. 1** Tahapan Proses KDD

Satu dari enam komponen paling penting dalam prosedur pencarian informasi ataupun pengetahuan dalam KDD (Yanto and Kesuma 2017) :

1. Bekerja dengan banyak data adalah langkah pertama.
2. Efisiensi diperlukan mengingat banyaknya data.

3. Mengutamakan akurasi.
4. Menuntut penggunaan terminologi yang canggih.
5. Memanfaatkan pembelajaran otomatis dalam beberapa hal.
6. Menghasilkan hasil yang menarik.

## **2.2 Datamining**

Data mining adalah proses mengumpulkan dan memanfaatkan Dengan menggunakan data historis, seseorang dapat mencari melalui kumpulan data yang sangat besar untuk mencari struktur, konsistensi, atau korelasi Pengambilan keputusan di masa depan dapat dibuat lebih baik dengan menggunakan temuan data mining (Tana, Marisa, and Wijaya 2018)

Berdasarkan (Hasugian 2020) Tujuan dari Data mining adalah proses mengeksplorasi dan menganalisis kumpulan data yang sangat besar untuk menemukan pola untuk mengekstrak informasi dan pengetahuan yang tepat dan mungkin berguna dari basis data yang sangat besar yang menarik.

Sesuai (Ikhwan 2018) Teknik pengumpulan informasi dari volume data yang sangat besar melalui penemuan pola dan korelasi yang tersembunyi dikenal sebagai data mining. Tindakan mengumpulkan data historis dan menerapkannya untuk mencari korelasi, tren, atau urutan dalam volume data yang sangat besar dikenal sebagai data mining, yang sering disebut sebagai database knowledge discovery (KDD). Pengambilan keputusan dapat dibantu oleh temuan data mining di masa depan. Pengenalan pola semakin jarang digunakan seiring dengan berkembangnya KDD karena dimasukkan ke dalam data mining.

Bidang interdisipliner datamining, yang mengintegrasikan Kesulitan yang terkait dengan mengekstraksi informasi dari kumpulan data yang sangat besar dapat diatasi dengan penggunaan pembelajaran mesin, pengenalan pola, statistik, basis data, dan alat visualisasi (Latifah 2018).

Satu atau lebih teknik pembelajaran mesin digunakan selama proses penggalian data untuk secara otomatis memeriksa dan mengekstrak pengetahuan. Metode pengembangan definisi dari ide-ide yang luas dengan melihat ilustrasi konkret dari ide yang harus dikuasai dikenal sebagai pembelajaran berbasis induksi, dan ini merupakan definisi lainnya. Penggalian data dengan menggunakan pendekatan ilmiah dikenal sebagai penemuan pengetahuan dalam basis data (KDD). Penggalian data merupakan salah satu fase dalam proses KDD dalam konteks ini (Sanjani, Fahmi, and Sindar 2019).

### **2.1.1 Pengelompokan Datamining**

Berdasarkan (Fajrin and Handoko 2018) Berdasarkan tugas yang dapat diselesaikan, data mining dikategorikan ke dalam berbagai kelompok, termasuk:

1. Deskripsi

Analisis dan peneliti terkadang hanya akan mencari bukti untuk mendukung pola dan tren yang mereka temukan. Sebagai contoh, mungkin saja para petugas jajak pendapat tidak dapat mendeteksi bahwa calon presiden yang kurang profesional akan menerima lebih sedikit dukungan. Diskusi tentang pola dan tren mencakup beberapa pembenaran untuk fenomena ini.

2. Estimasi

Variabel tujuan untuk estimasi adalah numerik dan bukan kategorikal, yang membedakannya dari klasifikasi. Catatan komprehensif yang memprediksi nilai variabel sasaran digunakan untuk membangun model. Nilai estimasi dari variabel target kemudian dihitung untuk tinjauan berikutnya tergantung pada prediksi yang dibuat oleh variabel. Sebagai contoh, kita dapat memperkirakan Tekanan darah sistolik pasien rumah sakit dihitung menggunakan usia, jenis kelamin, indeks massa tubuh, dan kadar garam darah. Model estimasi berdasarkan hubungan antara tekanan darah sistolik dan nilai variabel digunakan selama fase pembelajaran yang diantisipasi akan dikembangkan. Untuk kasus-kasus baru, seseorang dapat menerapkan model estimasi yang dihasilkan.

### 3. Prediksi.

Mirip terkait kategorisasi serta perkiraan, prediksi beda halnya dengan keduanya karena value hasilnya muncul di kemudian hari.

### 4. Klasifikasi

Ada variabel kategori target dalam kategorisasi. Misalnya, Penghasilan tinggi, penghasilan sedang, dan penghasilan rendah adalah tiga kategori yang diklasifikasikan ke dalam penghasilan.

### 5. Pengklusteran (*Clustering*)

Tindakan memisahkan data, observasi dan masalah ke dalam kategori-kategori item yang terkait dikenal sebagai pengelompokan. Sebuah klaster adalah sekelompok catatan yang memiliki kesamaan satu sama lain, tetapi tidak dengan klaster lainnya. Karena tidak ada variabel tujuan dalam pengelompokan, maka proses ini berbeda dengan klasifikasi. Tidak ada klasifikasi, estimasi, atau prediksi nilai variabel target yang dibuat selama proses pengelompokan. Algoritma

pengelompokan berusaha mempartisi data ke dalam kelas yang sama, memaksimalkan kesamaan pada sebuah kelompok sambil meminimalkan kemiripan dengan catatan dalam kelompok lain.

#### 6. Asosiasi

Dalam data mining, menemukan ciri-ciri yang sering muncul bersama adalah tujuan dari asosiasi. Hal ini lebih sering disebut sebagai analisis keranjang belanja di sektor korporat.

### **2.1.2 Tahapan – Tahapan Datamining**

Berdasarkan (Erwansyah 2019) satu dari sekian aplikasi pada seri Knowledge Discovery in Database (KDD) adalah datamining. KDD berfokus pada strategi penyatuan data serta identifikasi, analisis, dan penyajian pola dalam data menggunakan metode ilmiah. Berikut ini adalah langkah-langkah prosesnya:

#### 1. Pembersihan data

Prosedur ini bertujuan agar meniadakan noise dan data yang salah.

#### 2. Mengintegrasikan data

Dari banyak dasar adalah proses untuk melakukannya.

#### 3. Perubahan data

Transformasi data adalah prosedur yang digunakan untuk mengubah data ke dalam bentuk yang siap untuk ditambang.

#### 4. Penggunaan metode data mining

Metode untuk mengekstraksi pola dari data yang ada.

#### 5. Evaluasi pola yang ditemukan

Metode untuk mengubah pola menjadi informasi yang dapat diterapkan untuk membantu pengambilan keputusan.

## 6. Penyajian informasi

Presentasi pengetahuan dengan menggunakan metode grafis.

### 2.1.3 Manfaat Datamining

Sesuai (Erwansyah 2019) Terdapat dua perspektif dalam penggunaan data mining: perspektif bisnis dan perspektif ilmiah.

1. Data *mining* dapat digunakan di dunia korporat untuk mengendalikan peningkatan data secara eksponensial, dan metode komputasi dapat diterapkan untuk menawarkan informasi yang diperlukan, yang merupakan sumber daya yang dapat meningkatkan daya saing institusi.
2. Data *mining* dapat diterapkan secara ilmiah untuk mengumpulkan, mengevaluasi, dan menyimpan kumpulan data yang sangat besar dan real-time.

### 2.3 Metode *Datamining*

Tujuan dari pendekatan aturan asosiasi adalah untuk mengidentifikasi pola yang sering ada dalam sejumlah besar transaksi, yang masing-masing berisi sejumlah hal. Dengan mengidentifikasi pola di antara objek-objek ini, metode aturan asosiasi membantu sistem rekomendasi (Ikhwan 2018).

Menurut (Latifah 2018) Sebuah pendekatan datamining yang disebut aturan asosiasi digunakan untuk mengungkap hubungan yang unik atau korelasi yang menarik yang dapat disembunyikan oleh kumpulan data yang cukup besar. Sebuah pernyataan yang menyatakan bahwa jika X maka Y, di mana X adalah predikat dan Y adalah hasil, dikenal sebagai aturan asosiasi kumpulan item yang tidak akan pernah bisa digabungkan.

Menurut (Budiyasari et al. 2017) Pendekatan data mining untuk mengidentifikasi aturan dari kombinasi elemen adalah Penambangan aturan untuk asosiasi atau analisis asosiasi. di antara berbagai fase analisis asosiasi yang telah menarik minat banyak akademisi untuk mengembangkan algoritme yang efisien adalah analisis pola yang sering muncul (juga dikenal sebagai penambangan pola yang sering muncul). Dua metrik, yang disebut support dan confidence, dapat digunakan untuk menentukan signifikansi sebuah asosiasi.

1. Dukungan adalah proporsi dari gabungan item-item yang ada pada database.
2. Percaya diri mengukur seberapa kuat elemen-elemen dalam sebuah aturan asosiasi terkait satu sama lain.

Teknik dasar dari sebuah aturan asosiasi ialah (Budiyasari et al. 2017) :

- a. Analisis Pola Frekuensi Tinggi Algoritma Apriori

Langkah ini menentukan gabungan produk yang tergolong kriteria minimal nilai dukungan database. Rumus berikut ini dapat digunakan untuk menentukan nilai dukungan item:

$$Support A = \frac{\Sigma \text{Transaksi Mengandung A}}{\Sigma \text{Transaksi}} \times 100\%$$

**Sumber :** (Budiyasari et al. 2017)

**Rumus 2. 1** Perhitungan Support

- b. Kombinasi dua itemset pada tahap pembentukan Sedangkan rumus berikut ini digunakan untuk menentukan nilai support dari dua buah item:  $P(A,B)$   
 $= Support(A,B)$

$$Support\ A,B = \frac{\Sigma Transaksi\ Mengandung\ A\ dan\ B}{\Sigma Transaksi} \times 100\%$$

**Sumber :** (Budiyasari et al. 2017)

### **Rumus 2. 2** Perhitungan Support 2

Set item yang muncul lebih sering dari nilai minimum yang diberikan ( $\emptyset$ ) diklasifikasikan sebagai set item yang sering. Jika = 2, setiap itemset dengan frekuensi kemunculan lebih besar atau sama dengan 2 kali penyebutan. Fk adalah himpunan k-itemset yang sering muncul.

Mengatur Aturan Asosiasi Menghitung keyakinan hukum asosiasi A->B memungkinkan seseorang untuk mengidentifikasi aturan asosiasif dimana tergolong kriteria keyakinan minimal setelah semua pola frekuensi tinggi diidentifikasi. Rumus berikut ini digunakan untuk mendapatkan nilai keyakinan dari aturan A->B:

$$Confidence = \frac{\Sigma Transaksi\ Mengandung\ A\ dan\ B}{\Sigma Transaksi} \times 100\%$$

**Sumber :** (Budiyasari et al. 2017)

### **Rumus 2. 3** Perhitungan Confidence

Untuk dapat dipilih, aturan asosiasi harus disusun berdasarkan Support Confidence. Aturan N dengan hasil terbesar diterapkan sesering mungkin.

#### **2.3.1 Algoritma Apriori**

Untuk memastikan hubungan asosiatif antara sekelompok elemen, Metode pengambilan data berbasis asosiasi adalah algoritma apriori. Analisis pola frekuensi tinggi dari metode apriori mencari sekelompok objek dalam basis data yang memenuhi spesifikasi minimal value support (Yanto and Kesuma 2017).

Sesuai (Sikumbang 2018) Contoh aturan asosiasi pada data mining adalah algoritma apriori. Kelas ini juga mencakup Algoritma Berbasis Hasd dan teknik Induksi Aturan Umum selain Apriori. Menemukan aturan asosiatif antara sekelompok elemen dilakukan melalui Juga disebut sebagai pertambahan aturan asosiasi, analisis asosiasi. Analisis asosiasi adalah salah satu jenis penggalian data yang menjadi dasar bagi banyak metode Semua orang. Analisis pola frekuensi tinggi adalah salah satu tahap analisis asosiasi yang menarik minat untuk menciptakan algoritma yang efektif. Pentingnya sebuah aturan asosiasi dapat dipengaruhi oleh dua elemen: dukungan dan keyakinan. Sementara keyakinan adalah proporsi kombinasi item yang mungkin terjadi dalam database, dukungan adalah persentase dari kombinasi tersebut mengukur seberapa kuat item-item tersebut terkait satu sama lain dalam aturan asosiasi.

#### **2.4 Tanagra**

Salah satu alat penambangan data, Tanagra, menawarkan berbagai teknik data mining, yang mencakup pembelajaran mesin, pembelajaran statistik, dan sebagian besar penggalian data. Tanagra merupakan sebuah perangkat lunak gratis yang dapat digunakan oleh siapa saja selama mereka mematuhi persyaratan lisensi distribusi perangkat lunak dan memiliki akses ke kode sumber (Sikumbang 2018).

Menurut (Wahyudi, Bahri, and Handayani 2019) Perangkat lunak data mining bebas yang disebut Tanagra tersedia untuk digunakan dalam pendidikan dan penelitian. Perangkat lunak ini menyarankan sejumlah teknik penggalian data dari database, pembelajaran statistik, pembelajaran mesin, dan bidang analisis eksplorasi data. Tujuan dari proyek Tanagra ialah:

1. Menawarkan perangkat lunak penggalian data yang sesuai standar dan ramah pengguna kepada akademisi dan mahasiswa industri (terutama dalam hal desain dan penggunaan GUI) dan memungkinkan analisis data aktual dan fiktif.
2. Menyediakan sistem yang memudahkan para peneliti untuk mengintegrasikan teknik-teknik data mining mereka sendiri dan menilai seberapa baik teknik-teknik tersebut berfungsi.

## **2.5 Penelitian Terdahulu**

Implementasi Datamining untuk Memprediksi Metode Tingkat Penjualan Makanan Menggunakan Apriori merupakan judul penelitian tahun 2019 oleh Yuli Nur Indah Sari dkk. Hasil penerapan pendekatan algoritma Apriori pada data transaksi penjualan dengan aturan asosiasi dengan menggunakan program ini dapat membantu pihak manajemen dalam memperkirakan menu yang paling sering dipesan oleh konsumen, menemukan hubungan antar menu yang biasa dipesan bersamaan, dan meramalkan permintaan konsumen.

Dengan menggunakan algoritma Fp-growth, Ali Ikhwan (2018) melakukan penelitian dengan judul A novelty of datamining for promoting education. Telah ditunjukkan untuk menghasilkan aturan dengan menggunakan data sampel mahasiswa baru. Variabel data ketepatan FP-Growth yang dihasilkan dan persentase dalam mendefinisikan dukungan minuman dan kepercayaan diri minimal memiliki pengaruh besar terhadap set item yang sering muncul (frequent item set) yang dihubungkan untuk mengungkap variabel data yang akan digunakan sebagai strategi dalam mempromosikan pendidikan.

Penerapan Datamining dalam menganalisis tren penjualan di Swalayan 212 Mart di Lubuk Pakam Menggunakan Algoritma Apriori merupakan penelitian oleh Anzelia, dkk. yang diterbitkan pada tahun 2020. Temuan tersebut menunjukkan bahwa deteksi pola penjualan dengan menggunakan data transaksi penjualan memanfaatkan metode Apriori. untuk menetapkan aturan asosiasi yang dapat diterapkan pada konfigurasi tata letak dan taktik penjualan sesuai yang lebih baik, penelitian dilakukan dengan cara mengolah data pembelian barang oleh konsumen dan kemudian mencari hubungan antar barang yang dibeli.

Implementasi Datamining Penjualan Produk Pakaian Menggunakan Algoritma Apriori merupakan judul penelitian tahun 2019 oleh Sanjani, dkk. Diketahui bahwa pemilik toko dapat menggunakan aturan asosiatif yang dihasilkan oleh aplikasi untuk meningkatkan omzet penjualan, untuk menetapkan promosi harga untuk paket tertentu yang menggabungkan produk pakaian jadi yang sering ditawarkan bersama. Selain itu, pemilik toko dapat mengatur produk pakaian yang sering dibeli di rak-rak terdekat. Kombinasi pakaian yang sering dibeli bersama dapat ditemukan dengan menggunakan algoritma Apriori. digunakan oleh program untuk memberikan aturan asosiatif, nilai dukungan, dan nilai kepercayaan. Program ini dapat menampilkan fase-fase analisis algoritma Apriori, yang membantu dalam memahami bagaimana algoritma tersebut beroperasi.

Analisis Keranjang Belanja Algoritma Apriori di Mini Market Ayu merupakan penelitian tahun 2018 oleh Erlin Elisa. Minyak dan Susu memiliki dukungan dan kepercayaan paling tinggi, dengan nilai dukungan masing-masing 42,85% dan 85,71%. Diketahui bahwa penggunaan Algoritma Apriori dengan

data mining sangat berhasil dan dapat mempercepat proses pembuatan pola item set dari penjualan produk kebutuhan rumah tangga di Minimarket Ayu Tembesi-Batam.

Erma Delima Sikumbang (2018) melakukan penelitian dengan judul Aplikasi Datamining Penjualan Sepatu yang Dihitung Menggunakan Algoritma Apriori. Telah diketahui bahwa Membuat korelasi antara frekuensi penjualan sepatu dan permintaan konsumen membutuhkan penggunaan data mining dan algoritma apriori. pelanggan. Algoritma apriori membantu dalam perumusan taktik penjualan sepatu. Merek sepatu yang paling populer, menurut penelitian, adalah New Balance (91,67%), Adidas (75%), Geox (50%), Nike (41,67%), dan Palladium (41,67%).

Penelitian Kamil Erwansyah, Implementasi Datamining untuk Menganalisis Sebuah penelitian tentang korelasi antara persediaan stok dan data pendapatan untuk produk kimia di PT. Grand Multi Chemicals dengan menggunakan Algoritma FP (Frequent Pattern) Development telah dipublikasikan pada tahun 2019. Tren pembelian barang kimia untuk memperoleh Dengan menggunakan metode penggalian data untuk memeriksa hubungan antara data penjualan biologis dan tingkat persediaan, dimungkinkan untuk mengidentifikasi senyawa yang layak untuk diunduh pada saat yang sama untuk menjamin ketersediaan yang cukup untuk barang yang sesuai dan untuk meminimalkan pembelian senyawa yang jarang digunakan di PT. Grand Multi Chemicals.

Datamining Di terminal tiket tur dan perjalanan, algoritme apriori digunakan untuk menjual tiket pesawat Batam merupakan penelitian yang

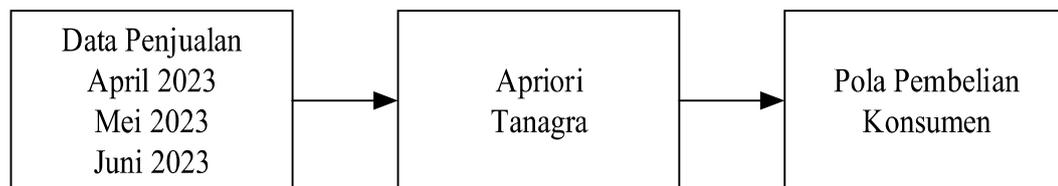
dilakukan oleh Narti Eka Putria (2018). Di dalam Batam Tour & Travel Ticket Terminal, di mana metode apriori datamining telah diidentifikasi untuk mengungkap tren konjungsi itemset dan peraturan asosiasi, Lion dan Sriwijaya serta Garuda dan Sriwijaya memiliki nilai dukungan dan kepercayaan terbesar, dengan nilai persetujuan 50% dan keandalan 70%. Menurut informasi berikut, Sriwijaya akan bertindak demi kepentingan terbaik Anda jika Anda membeli tiket Lion, dengan nilai bantuan 50% dan tingkat kepercayaan 70%, dan jika orang membeli tiket Garuda, hal yang sebaliknya akan terjadi.

Penerapan Metode Datamining Mateus Paga Tana dan rekan-rekannya mempublikasikan sebuah penelitian pada tahun 2018 yang berjudul Evaluasi Keranjang Pasar berdasarkan Statistik Pembelian Barang di Toko Oase. dengan memanfaatkan algoritma Apriori. Hubungan antara satu hal dengan hal lainnya terjalin dengan baik. Dan digunakan untuk mengontrol posisi komoditas berdasarkan keterkaitan tersebut. Nilai Di mana barang yang baik harus diletakkan dapat diputuskan dengan menggunakan nilai dukungan dan kepercayaan. Barang-barang ini dicantumkan di bagian atas dan bawah daftar karena merupakan barang yang paling populer di kalangan pelanggan. Karena ada kemungkinan besar bahwa dua atau lebih barang akan dibeli bersamaan, barang dengan tingkat kepercayaan yang sangat baik ditempatkan berdekatan. Tingkat kepercayaan dan dukungan tertinggi ditunjukkan untuk produk tembakau, makanan espresso, pasta yang dimasak, dan sayuran berminyak; sebagai hasilnya, penerapan algoritme Apriori dalam pendekatan penggalian data sangat efisien dan dapat mempercepat pembuatan tren kombinasi campuran itemset dari penjualan barang dagangan di OASE Store. Produk dagangan diposisikan berdekatan di dalam toko.

Penerapan datamining menggunakan pendekatan Yori Apridonal M, dkk. (2019) melakukan penelitian tentang penggunaan aturan asosiasi dan algoritma Apriori untuk menilai tren penjualan produk. Diketahui bahwa, dengan nilai support minimum 40% dan tingkat kepercayaan minimum 60%, algoritma Apriori dapat menghasilkan hingga 6 aturan asosiasi.

Judul penelitian tahun 2020 oleh Aji Setiawan dkk. adalah Analisis Keranjang Pasar Busana Muslim Trendi pada E-Commerce Shop Algoritma Apriori. Algoritma Apriori dikenal untuk membantu klien dalam membuat keputusan pembelian dengan meninjau halaman saran produk dengan cepat. Pendekatan ini dapat membantu pelanggan dalam memilih barang selain membantu mereka dalam pemilihan produk administrator dalam mempelajari statistik transaksi pembelian pelanggan.

## 2.6 Kerangka Pemikiran



**Sumber :** (Peneliti 2023)

**Gambar 2. 2** Kerangka Penelitian