

BAB II

TINJAUAN PUSTAKA

2.1. Teori Dasar

Teori-teori yang menjadi landasan penelitian ini akan dijelaskan peneliti pada bab ini, yaitu sebagai berikut:

2.1.1. *Clustering*

Analisis *cluster* adalah proses membagi data menjadi kelompok. Berbeda dengan metode klasifikasi, kelasnya adalah tidak ditentukan sebelumnya. Kelompok dan *cluster* yang datanya akan dibagi, dan bahkan berapa banyak kelompok yang berbeda akan dibagi menurut kesamaan data yang tersedia. Analisis *cluster* digunakan dalam banyak bidang seperti biologi, kedokteran, antropologi dan pemasaran (Kusak, Unel, Alptekin, Celik, & Yakar, 2021). Pada penelitiannya (Kusak et al., 2021) menggunakan algoritma *K-Means* untuk mengklasifikasikan wilayah sebelum terjadinya tanah longsor berdasarkan faktor lingkungan seperti kepadatan vegetasi, kecuraman lereng, dan jarak dari garis aliran air. Masing-masing dari 5 klaster yang dibuat dengan metode ini memiliki profil risiko tanah longsor yang berbeda. Berdasarkan salah satu *cluster* yang ditemukan, 38% wilayahnya berisiko longsor karena kemiringan lereng yang curam, vegetasi yang jarang, dan lokasi yang jauh dari aliran air.

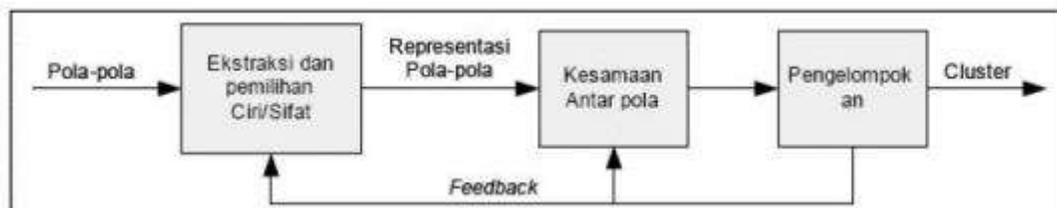
Menurut (Indraputra & Fitriana, 2020) *clustering* adalah proses membagi suatu kumpulan data menjadi beberapa subset sehingga data pada setiap subset mempunyai arti yang bermakna, atau lebih spesifik lagi, mengelompokkan objek-objek yang sebanding ke dalam kelompok-kelompok yang berbeda. Mengorganisasikan data ke dalam beberapa grup atau *cluster* di mana *item* dalam satu grup memiliki sifat yang sama dengan objek lain dalam grup tersebut adalah tujuan utama *clustering*. Penerapan teknik ini meliputi identifikasi pola, analisis pola konsumen, dan segmentasi pasar. Algoritma pengelompokan (*clustering*) dibagi menjadi dua bagian yaitu, secara partisi dan hierarki. Algoritma partisi memutuskan semua kelompok pada waktu tertentu, sedangkan algoritma hierarki menempatkan klaster secara berurutan ketika klaster ditentukan sebelumnya. Pada penelitiannya (Indraputra & Fitriana, 2020) mengekstrak dan mengolah data COVID-19 dari *Kaggle* menggunakan teknik *data mining K-Means clustering*. Data dikelompokkan berdasarkan *mean* terdekat, sehingga memungkinkan wilayah untuk dikelompokkan ke dalam *cluster* sesuai dengan dampak pandemi COVID-19 terhadap wilayah tersebut. Informasi *cluster* kemudian dapat digunakan untuk menentukan prioritas dan memandu bantuan COVID, memastikan bahwa pandemi COVID-19 ditangani seefektif mungkin. Dua *cluster* data dihasilkan dari temuan pengolahan data, klaster 2 memiliki lebih banyak penyakit dan kematian dibandingkan klaster 1, oleh karena itu wilayah penanganan klaster tersebut harus menjadi prioritas utama.

Secara umum, metode partisi mengasumsikan bahwa kumpulan data dapat direpresentasikan oleh prototipe *cluster* terbatas dengan miliknya sendiri. Oleh karena

itu, mendefinisikan ketidaksamaan (atau jarak) antara titik dan prototipe *cluster* sangat penting untuk metode partisi. Diketahui algoritma *k-means* adalah metode partisi tertua dan populer. *K-means clustering* telah banyak dipelajari dengan berbagai macam perluasan dalam literatur dan diterapkan di berbagai bidang substantif. Namun, *k-means* ini algoritma pengelompokan biasanya dipengaruhi oleh inisialisasi dan perlu diberikan sejumlah *cluster* terlebih dahulu (Sinaga & Yang, 2020). Penelitian oleh (Sinaga & Yang, 2020) berupaya untuk mengatasi kekurangan teknik *K-Means* konvensional, yang bergantung pada inisialisasi awal dan penentuan jumlah *cluster*. Tanpa memerlukan pengaturan atau parameter tertentu, peneliti menciptakan algoritma *Unsupervised K-Means (U-K-means)*, yang secara otomatis dapat mengidentifikasi jumlah *cluster* yang ideal. Pendekatan ini dimaksudkan untuk meningkatkan efisiensi pengelompokan data di berbagai aplikasi yang lebih luas sekaligus menurunkan kompleksitas komputasi. Menurut hasil penelitian, *U-K-means* mengungguli teknik eksperimen lain dan versi *K-Means* konvensional dalam mengidentifikasi *cluster*. Pendekatan ini serbaguna untuk banyak domain, termasuk pengenalan pola dan pemrosesan data besar, dan sangat berguna untuk data dengan pola rumit atau ketika jumlah kluster ideal tidak diketahui sebelumnya.

Menurut (Aulia, 2021) *cluster* adalah kumpulan objek data terkait yang dikelompokkan bersama dalam *cluster* yang sama dan dikontraskan dengan objek data dari *cluster* lainnya. Pengelompokan suatu benda ke dalam satu atau lebih *cluster* akan menghasilkan tingkat kemiripan yang tinggi antara objek-objek yang ada dalam *cluster* tersebut dengan objek yang lain.

Dalam (Zuhal, 2022) dijelaskan bahwa terdapat 2 kategori teknik *clustering*, yaitu *Partitioning clustering* dan *Hierarchical clustering*. Langkah pertama dalam metode pengelompokan data pengelompokan partisi (*Partitioning clustering*) adalah menentukan terlebih dahulu berapa banyak *cluster* yang diinginkan (dua *cluster*, tiga *cluster*, dst). Setelah jumlah *cluster* telah ditentukan, prosedur *cluster* dijalankan tanpa mengikuti proses hierarki. Hierarki *clustering* merupakan teknik pengelompokan data yang dimulai dengan menggabungkan dua atau lebih hal yang paling sebanding atau dekat kesamaannya. Prosedur tersebut kemudian dilakukan untuk objek kedua yang terdekat. Dengan demikian, *cluster* akan menyerupai pohon dengan hierarki berbeda antar komponen yang disusun dari yang paling mirip hingga yang paling tidak mirip. Berikut adalah tahapan pada proses *clustering*:



Gambar 2. 1 Tahapan Proses Clustering

Sumber : (Zuhal, 2022)

2.1.2. *Association Rules*

Salah satu tugas *data mining* deskriptif yang berfokus pada menemukan aturan asosiasi antar objek data adalah aturan asosiasi (*association rules*). Menentukan seberapa sering kombinasi objek muncul dalam *database* dikenal sebagai pola yang

sering berulang adalah langkah utama dalam proses aturan asosiasi (Alinafiah, Octariadi, Informatika, & Teknik, 2018).

Seperti dalam analisis transaksi ritel atau mesin rekomendasi, pendekatan Association rules biasa disebut sebagai analisis keranjang pasar. Analisis keranjang dengan mengidentifikasi hubungan antara barang-barang yang sering dibeli, pasar dapat dianggap sebagai proses analisis pola pembelian konsumen. Tujuan dari penelitian ini adalah untuk mengungkap pola tersembunyi dari kemunculan item ke dalam bentuk aturan yang mudah diidentifikasi, bukan untuk mengantisipasi bagaimana item akan muncul dengan menggunakan teknik seperti klasifikasi atau regresi. *Antecedent* dan *consequent* adalah istilah yang digunakan untuk menggambarkan informasi yang dihasilkan dalam bentuk hubungan “*if-then*” atau “jika-maka”. Bagian “jika” diwakili oleh *antecedent*, dan bagian “maka” diwakili oleh *consequent*. Membuat aturan *Association Rules* adalah $A \rightarrow B$, dimana item B adalah konsekuensinya dan item A adalah antesedennya. Aturan $A \rightarrow B$ misalnya menyatakan bahwa dalam transaksi yang melibatkan belanja konsumen, jika pelanggan membeli barang A maka ia juga harus membeli barang B (Setiawan, Sedyono, & Sembiring, 2020).

Menurut (Suseno, 2024) hubungan atau asosiasi penting antara sekelompok objek atau kualitas dalam data dapat ditemukan dengan menggunakan teknik aturan asosiasi (*Association rule*) dalam *data mining*. Metode ini biasanya digunakan dalam analisis keranjang belanja untuk mengidentifikasi pola pembelian yang berulang. Algoritma apriori adalah salah satu dari beberapa kemungkinan algoritma yang

digunakan pada *Association rule*. Algoritma dalam penambangan data yaitu, teknik apriori digunakan untuk mengidentifikasi pola hubungan antar objek dalam kumpulan data transaksi. Hubungan antara objek-objek yang biasanya muncul bersamaan dalam transaksi diidentifikasi oleh aturan asosiasi ini. Nilai *support* minimal dan nilai keyakinan minimum merupakan dua syarat untuk menggunakan algoritma apriori. Persentase atau jumlah minimum transaksi di mana suatu aturan harus terjadi agar dianggap signifikan ditunjukkan oleh nilai dukungan minimum. Nilai keyakinan minimum, sebaliknya, menetapkan proporsi minimal transaksi di mana elemen-elemen terkait harus muncul bersama-sama dalam suatu transaksi agar aturan tersebut dianggap relevan atau patut untuk diperiksa lebih lanjut.

Menurut (Oktaviani, 2024) salah satu teknik penambangan data untuk mengidentifikasi tren frekuensi tinggi dalam sejumlah besar transaksi adalah aturan asosiasi (*association rules*). Pengguna harus memasukkan nomor *min confident* dan *min support* yang telah ditentukan sebelumnya untuk mendapatkan tingkat frekuensi. Terdapat dua metrik, yaitu *support* dan *confident* yang digunakan untuk menilai aturan asosiasi (Fuadi, Bhakti, & Premana, 2024).

1. *Support*: Ini menampilkan frekuensi terjadinya *item* atau kelompok hal tertentu dalam setiap transaksi. Kesesuaian suatu *item* atau sekelompok item sebagai dasar perhitungan ditentukan oleh ukurannya *confidence*.
2. *Confidence*: Sejauh mana 2 *item* mempunyai hubungan bersyarat satu sama lain ditunjukkan oleh *confidence*. Ini mengukur seberapa sering *item* atau *itemset* B muncul ketika *itemset* A juga muncul.

Kedua metrik ini memungkinkan peneliti dapat menilai signifikansi atau kepentingan aturan asosiasi dalam penambangan data.

Menurut (Setiawan et al., 2020) tahapan utama pada metode analisis asosiasi terdiri dari 2 tahap yaitu:

1. Menganalisis Pola Frekuensi Tinggi

Pada tahap ini, database mencari kombinasi *item* yang memenuhi persyaratan nilai *support* minimum. Rumus berikut digunakan untuk menentukan nilai *support* dari 1 *item*:

$$Support (A) = \frac{\sum \text{Transaksi Item A}}{\sum \text{Transaksi}}$$

Rumus 2. 1 *Support Itemset-1*

Persamaan nilai *support* penampilan 1 *item* dengan membagi jumlah transaksi yang mengandung *item* A dengan jumlah total transaksi, maka dukungan nilai untuk *item* A dapat dihitung.

Sementara itu, gunakan rumus berikut untuk mencari nilai *support* dari 2 *item*:

$$Support (A \rightarrow B) = \frac{\sum \text{Transaksi Item A dan B}}{\sum \text{Transaksi}}$$

Rumus 2. 2 *Support Itemset-2*

Banyaknya transaksi yang mengandung *item* A dan B dibagi dengan jumlah total transaksi merupakan persamaan munculnya nilai *support item* A dan *item* B.

2. Pembentukan Aturan Asosiasi

Setelah keberhasilan identifikasi setiap pola frekuensi tinggi, maka aturan asosiatif terbentuk. Pada proses ini akan menghitung keyakinan (*confidence*) aturan asosiatif untuk menemukan aturan yang memenuhi kondisi minimal nilai *confidence*. Dalam aturan asosiasi, *confidence* (nilai kepastian) mengacu pada asosiasi antar *item* yang kuat (Aulia Miranda, Fahrullah, & Kurniawan, 2022) Rumus berikut digunakan untuk menentukan nilai keyakinan (*confidence*) yaitu:

$$Confidence (B / A) = \frac{\sum \text{Transaksi Item A dan B}}{\sum \text{Transaksi Item A}}$$

Rumus 2.3 *Confidence*

2.1.3 *K-Means*

Salah satu metode tradisional untuk menyelesaikan masalah *clustering* adalah algoritma *K-Means* (Syamfithriani, Mirantika, & Trisudarmo, 2023). *Clustering* adalah teknik untuk mengidentifikasi dan mengklasifikasikan data saat ini berdasarkan kesamaan antara kumpulan data yang satu dengan kumpulan data lainnya. Salah satu teknik penambangan data tanpa pengawasan adalah pengelompokan, yang dapat digunakan tanpa bimbingan atau pelatihan dan tidak memerlukan tujuan keluaran (*output*) (Adani et al., 2019). Pada penelitian yang dilakukan oleh (Syamfithriani et al., 2023) bertujuan untuk mencegah dan mengendalikan penyebaran penyakit diare pada balita di Kabupaten Kuningan, penelitian ini bermaksud menyusun rencana pemetaan wilayah prioritas. Metode yang digunakan adalah dengan membandingkan dua

algoritma, yaitu algoritma *K-Means* dan algoritma *K-Medoids* melalui pendekatan *data mining clustering*. Berdasarkan hasil pendekatan ini, jumlah *cluster* optimal pada algoritma *K-Means* adalah 3, sedangkan pada algoritma *K-Medoids* adalah 2. Hal ini menunjukkan keunggulan algoritma *K-Means* dibandingkan pendekatan *K-Medoids*.

Menurut (Aulia, 2021) metode data *K-Means* merupakan teknik pengelompokan non-hierarki yang bertujuan untuk membagi data menjadi satu atau lebih *cluster* atau kelompok sehingga data yang mempunyai karakteristik serupa dikelompokkan menjadi satu *cluster*, dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok yang berbeda. Untuk mengevaluasi data yang kompleks dan mengelompokkan *item* data menjadi banyak kelompok atau *cluster* berdasarkan karakteristik yang sebanding, *K-means clustering* adalah teknik *data mining* yang mengelompokkan data berdasarkan titik pusat *cluster* (*centroid*) (Laksono, Syahidin, & Yunengsih, 2024). Dalam penelitiannya (Aulia, 2021) menggunakan metode *K-Means* untuk mengelompokkan penjualan pestisida tanaman padi selama satu musim. *Cluster* dibagi menjadi tiga kategori, yaitu sangat laris, laris, dan tidak laris. Teridentifikasi 53 produk pestisida yang sangat laris, 21 produk yang teridentifikasi laris, dan 126 produk yang teridentifikasi menjual pestisida yang kurang laris.

Menurut (Fajar, Rahaningsih, Dana, Informatika, & Akuntansi, 2024) salah satu pendekatan *data mining* yang menggunakan teknik *clustering* berbasis jarak divisi adalah algoritma *K-Means*. Keuntungan menggunakan *K-Means* mencakup kemudahan pemahaman dan implementasi, serta dampak pengelompokan yang kuat,

yang menjadikannya pilihan populer untuk penelitian lapangan. Teknik pengelompokan data disebut analisis kluster. Kelas-kelas tersebut tidak diatur sebelumnya, berbeda dengan proses klasifikasi. Kesamaan data tertentu menentukan kelompok dan kluster mana yang akan dipecah menjadi data, serta berapa banyak kelompok berbeda yang akan memisahkan data tersebut. Banyak disiplin ilmu, termasuk biologi, kedokteran, antropologi, dan pemasaran, menggunakan analisis kluster. Penelitian oleh (Fajar et al., 2024) data penjualan obat di Apotek An-Naafi dianalisis dengan menggunakan pendekatan *K-Means*. Data obat dikategorikan berdasarkan tingkat penjualan (tinggi dan rendah) untuk mengetahui tren dan pola yang memberikan informasi berharga bagi apotek. Dua *cluster* utama yang mewakili tingkat penjualan rendah dan tinggi dibentuk dengan menerapkan pendekatan *K-Means* dengan inisialisasi *cluster* yang tepat. Karakteristik penjualan obat yang berbeda-beda ditampilkan oleh setiap *cluster*, menawarkan informasi tentang preferensi konsumen dan kemungkinan penjualan yang lebih tinggi. *K-Means* secara efektif digunakan untuk membuat *cluster*, menawarkan informasi mendalam untuk meningkatkan produktivitas di sektor farmasi. Hasil *cluster* dipengaruhi oleh penggunaan jarak *Euclidean*, *cluster* 0 lebih beragam, sedangkan *cluster* 1 menonjol karena memiliki penjualan dan pendapatan yang besar.

Pendekatan paling populer yang menggunakan kriteria *square-error* adalah teknik pengelompokan partisi *K-means*. MacQueen adalah orang pertama yang menggunakan *K-means*. Menemukan partisi yang meminimalkan kesalahan kuadrat total untuk sejumlah *cluster* adalah tujuan keseluruhannya (Kusak et al., 2021). Dalam

(Efran, Khairil, & Jumadi, 2022) dikatakan bahwa model *centroid* adalah teknik atau metode yang dilakukan *clustering K-Means*. Model *centroid* adalah model yang menggunakan *centroid* untuk membentuk *cluster*. *Centroid* adalah nilai tengah *cluster*, dan *centroid* digunakan untuk menentukan seberapa jauh suatu objek dari *centroid*. Jika suatu objek data paling dekat dengan *centroid cluster*, maka dianggap sebagai anggota *cluster*. Dalam (Fajar et al., 2024) segmentasi gambar digital dan identifikasi objek dapat dilakukan dengan menerapkan pendekatan *K-Means Clustering*. Segmentasi dan identifikasi dilakukan dengan membandingkan data pengujian yang ingin diidentifikasi dengan data pelatihan yang dimasukkan sebelumnya. Skor kesamaan berdasarkan jumlah *cluster* pada gambar yang dilatih sebelumnya adalah hasil akhir dari penelitian ini.

Menurut (Fitri et al., 2023) langkah-langkah pada algoritma *k-means* adalah sebagai berikut:

1. Menentukan *cluster* (k) yang akan dibentuk dalam dataset
2. Mengatur nilai pusat (*Centroid*)

Nilai *centroid* awalnya ditentukan secara acak dan kemudian diiterasi menggunakan rumus berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Rumus 2. 4 Centroid

Keterangan:

- a. Untuk variabel ke-j, V_{ij} adalah rata-rata *centroid* dari *cluster* ke-I.

- b. Indeks *cluster* adalah i dan k .
 - c. N_i adalah jumlah anggota *cluster* ke- i .
 - d. Indeks variabel dilambangkan dengan j .
 - e. Untuk *cluster* tersebut, X_{kj} adalah nilai data ke- k dari variabel ke- j .
3. Untuk setiap *record*, jarak terdekat ke *centroid* dihitung dengan menggunakan rumus Jarak *Euclidean* (*Euclidean Distance*) sebagai berikut:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Rumus 2. 5 *Euclidean Distance*

Keterangan:

- a. D_e melambangkan jarak *Euclidean*.
 - b. I melambangkan jumlah benda atau objek
 - c. Koordinat benda adalah x, y .
 - d. Koordinat pusat (*centroid*) adalah s, t .
4. Mengelompokkan objek berdasarkan seberapa dekat benda-benda tersebut dengan *centroid*.
5. Hingga nilai ideal tercapai, iterasi dari langkah 3 hingga 4 dilakukan berulang kali.

2.1.4 Algoritma Apriori

Menurut (Liu, Wang, Miao, & Ren, 2022) salah satu algoritma yang paling banyak digunakan dalam *data mining* aturan asosiasi adalah algoritma apriori. Aturan asosiasi yang dihasilkan dengan metode ini adalah aturan asosiasi tipe *Boolean* tingkat tunggal dalam hal klasifikasi, seperti yang ditemukan dalam sistem komersial dan medis. Algoritma aturan asosiasi apriori secara berulang mencari basis data transaksi lapis demi lapis untuk menemukan kelompok *item* umum. Dalam penelitian (Liu et al., 2022) Algoritma Apriori digunakan dalam mengatasi beberapa masalah, seperti kebutuhan untuk memindai database transaksi kasus beberapa kali, menghasilkan kumpulan data yang besar dan aturan yang berlebihan. Untuk mengatasi masalah tersebut, diusulkan algoritma penambahan data aturan asosiasi yang menggabungkan matriks pengelompokan dan strategi pemangkasan, yang meningkatkan algoritma dengan menggunakan metode matriks pengelompokan untuk mengompresi database transaksi yang disimpan dan memperkenalkan metode strategi pra-pemangkasan dan pasca-pemangkasan berdasarkan menambahkan kondisi kendala. Hasil penelitian menunjukkan bahwa jumlah himpunan kandidat yang dihasilkan oleh algoritma Apriori yang dioptimalkan dalam kondisi ambang batas dukungan yang berbeda lebih sedikit dibandingkan dengan algoritma tradisional, dan waktu eksekusi algoritma tersebut sangat berkurang, sehingga meningkatkan kinerja efisiensi eksekusi algoritma

Dengan membangun hubungan antara data penjualan, produk algoritma apriori dapat digunakan untuk menghasilkan ide untuk strategi penjualan di masa depan.

Konsumen akan menerima apa pun yang dipesan berdasarkan tren pembelian konsumen. Pada penelitian (Aulia Miranda et al., 2022) terjadi masalah yaitu selama dua tahun jumlah transaksi penjualan di Toko semakin meningkat, namun data dari transaksi tersebut hanya digunakan untuk laporan dan tidak digunakan untuk menentukan rencana bisnis. Data transaksi tersebut dapat digunakan untuk melihat bagaimana setiap jenis produk yang dibeli konsumen secara bersamaan berhubungan dengan produk lainnya. Metode algoritma apriori digunakan dalam penelitian ini untuk menganalisis data karena algoritma apriori mampu mengelola proses penambahan data dalam jumlah besar secara efisien. Berdasarkan temuan hasil penelitian, dihasilkan 6 (enam) aturan asosiasi dengan nilai keyakinan berkisar antara 58% hingga 75% dari 568 data penjualan Sheshop dengan menggunakan nilai dukungan minimal 10% dan nilai keyakinan minimal 50%. Seluruh aturan mempunyai tingkat korelasi positif. Dengan nilai kepercayaan 75%, dihasilkan dua produk yaitu Al-Quran dan tasbih yang sering dibeli bersamaan berdasarkan 6 (enam) aturan asosiasi yang ditemukan.

Berdasarkan (Nainggolan & Elisa, 2022) Algoritma apriori merupakan algoritma pengambilan yang menentukan hubungan asosiasi suatu kombinasi elemen dengan menggunakan aturan asosiasi (*association rule*). Dengan menggunakan metode resolusi apriori, sampel frekuensi tinggi dianalisis untuk menemukan kombinasi objek yang memenuhi persyaratan nilai dukungan minimum statistik dasar. Algoritma yang dikenal sebagai algoritma apriori adalah algoritma yang menggunakan aturan asosiasi untuk mengambil data dan memastikan hubungan antara sekelompok elemen. Dalam penelitian (Nainggolan & Elisa, 2022) permasalahan yang terjadi yaitu data transaksi

penjualan yang tidak dikelola pada toko Surya Elektronik mengakibatkan pola pengadaan produk yang tidak tepat dan tata letak yang tidak mengikuti perilaku pelanggan. Tingkat strategi penjualan dan pemasaran sangat dipengaruhi oleh hal tersebut. Oleh karena itu, untuk memastikan persediaan barang sesuai dengan kebutuhan pelanggan, diperlukan suatu analisis yang dapat membantu Toko Surya Elektronik dalam memberikan informasi tentang perilaku konsumen dan barang apa saja yang dibutuhkan. Penelitian tersebut menggunakan software Tanagra untuk mengelola teknik komputasi *data mining* yang menggunakan algoritma Apriori. Berdasarkan nilai dukungan (*support*) dan kepercayaan (*confidence*), penelitian ini akan membantu Toko Surya Elektronik dalam mengatur persediaan dan tata letak produk sesuai dengan minat pelanggan. Analisis telah menunjukkan bahwa hasil penggunaan pengumpulan data asosiasi untuk mengungkap informasi penjualan baru dapat digunakan sebagai perkiraan untuk strategi pemasaran dan penjualan di masa depan yang bertujuan untuk meningkatkan pendapatan dan mempertahankan pelanggan.

Dalam (Rifania, Saniman, & Azlan, 2023) salah satu jenis aturan asosiasi yang digunakan dalam *data mining* adalah algoritma Apriori. Analisis afinitas dan analisis keranjang pasar adalah istilah yang sering digunakan untuk menggambarkan aturan yang menunjukkan hubungan dalam atribut tertentu. Banyak peneliti menggunakan analisis asosiasi atau penambangan aturan asosiasi sebagai teknik penambangan data untuk membuat algoritma yang efektif. Terdapat dua kriteria yaitu dukungan (*support*) dan keyakinan (*confidence*), dapat digunakan untuk menentukan apakah suatu

hubungan signifikan atau tidak. *Confidence* (nilai kepastian) adalah derajat korelasi antar *item* dalam aturan asosiasi, dan *support* (nilai dukungan) adalah persentase *item* tersebut dalam *database*. Algoritma apriori menggunakan teknik pengambilan dan didasarkan pada aturan asosiasi. Data menentukan hubungan asosiasi untuk kombinasi *itemset* dengan menggunakan teknik aturan asosiatif.

Dalam (Saefudin & Fernando, 2020) dijelaskan bahwa metode sederhana untuk mengidentifikasi kumpulan item umum pada aturan asosiasi *Boolean* adalah algoritma apriori. Konsep utama dari algoritma apriori adalah:

1. Mencari database transaksi untuk *frequentitemset*, yang merupakan kumpulan *item-item* dengan dukungan (*support*) minimal.
2. Menghapus *itemset* frekuensi rendah sesuai dengan tingkat dukungan minimal yang telah ditetapkan sebelumnya.

Membuat aturan untuk asosiasi antar *itemset* yang memenuhi keyakinan nilai *minimal database* adalah langkah berikutnya. Algoritma apriori menggunakan dua langkah utama untuk membuat kandidat *itemset*, yaitu sebagai berikut:

1. Langkah penggabungan (*JoinStep*)
Setiap *item* digabungkan dengan *item* lain selama prosedur hingga tidak ada lagi kombinasi yang dapat dibuat.
2. Pemangkasan (*PruneStep*)
Tahapan ini melibatkan penggabungan temuan *item* dan kemudian memangkasnya sesuai dengan aturan yang disebutkan sebelumnya.

2.1.5 Pola Pembelian

Menurut (Adha, 2022) faktor-faktor yang mempengaruhi perilaku pelanggan yaitu ada 5 fase perilaku konsumen yang terjadi selama proses pembelian yakni identifikasi kebutuhan, pencarian informasi, evaluasi informasi, pembelian, dan pasca pembelian. Budaya adalah faktor paling mendasar yang mempengaruhi perilaku dan keinginan, dan budaya mempunyai dampak yang mendalam dan luas terhadap perilaku konsumen. Elemen budaya ini mencakup hal-hal seperti kelas sosial ekonomi, budaya, dan subkultur komponen sosial. Beberapa contoh faktor sosial yang mempengaruhi perilaku konsumen adalah peran dan status sosial konsumen, keluarga mereka, dan kelompok referensi mereka. Keputusan pembelian pelanggan juga dipengaruhi oleh kepribadiannya. Faktor-faktor tersebut antara lain jenis, jenis kelamin, usia, pekerjaan, gaya hidup, keadaan keuangan, kepribadian, dan konsep diri konsumen.

Menurut (Tonda, Muh. Raditya Hanif F, & Tuhu Setya Ning Tyas, 2022) proses di mana pelanggan melakukan seluruh tindakan mereka, termasuk mencari, meneliti, dan mengevaluasi suatu produk sebelum membelinya, dikenal sebagai perilaku konsumen. Faktor paling mendasar yang mempengaruhi keputusan konsumen untuk membeli atau tidak adalah perilakunya. Studi tentang perilaku konsumen melihat bagaimana orang, kelompok, dan organisasi memilih, membeli, menggunakan, dan menilai barang untuk memenuhi kebutuhan dan keinginan mereka.

Perilaku pembelian seseorang dapat dipengaruhi oleh berbagai macam faktor, seperti kualitas barang, harga yang ditawarkan, cara yang dipromosikan, kemudahan

pencairan di tempat pembelian, dan masih banyak lagi. Dengan demikian, organisasi komersial atau korporasi dapat mengembangkan konsep rencana pemasaran yang akan digunakan untuk menarik pelanggan potensial baru dan mempertahankan pelanggan setia yang sudah ada dengan meneliti perilaku konsumen. Produsen akan memperoleh pemahaman yang lebih baik tentang strategi pemasaran dan komunikasi jika mereka memahami elemen-elemen yang mempengaruhi keputusan dan perilaku konsumen. Selain itu, produsen harus memahami perilaku konsumen dan proses pengambilan keputusan. Lebih detail strategi pemasaran mengenai target pasar selalu dapat direncanakan oleh produsen. Masalah sering kali muncul ketika produsen menghadapi berbagai jenis perilaku konsumen ketika mereka melakukan pembelian yang tidak masuk akal (Acai Sudirman, Syafika Alaydrus, Siti Rosmayati, Lucky Nugroho, Opan Arifudin, Ita Musfirowati Hanika, & Anne Haerany, Fenny Damayanti Rusmana, n.d.).

Dalam (Nuryani, Pattiwael, & Iqbal, 2022) menyatakan bahwa faktor pertama yang mempengaruhi pembelian adalah adanya promosi. Secara umum, promosi adalah serangkaian tindakan yang digunakan untuk membujuk pelanggan agar melakukan pembelian. Selain itu, promosi tidak selalu digunakan sebagai sarana komunikasi antara pembeli dan penjual, sebaliknya mereka juga digunakan sebagai sarana bagi penjual untuk mempengaruhi keputusan pembeli dalam melakukan pembelian. Apabila promosi berhasil membangkitkan minat beli konsumen, maka pada akhirnya calon konsumen akan ingin memperoleh suatu produk tertentu.

Penelitian (Anam, 2024) menyatakan bahwa keputusan pembelian konsumen pada platform *e-commerce* sangat dipengaruhi oleh variabel ekonomi. Dengan

memberikan insentif kepada pelanggan untuk melakukan pembelian, kebijakan harga yang kompetitif dan promosi yang sukses dapat mempengaruhi pilihan konsumen. Pelanggan biasanya lebih cenderung membeli ketika mereka menganggap harga lebih menarik atau ketika mereka mendapat keuntungan dari promosi. Namun harga bukanlah satu-satunya pertimbangan ekonomi. Penelitian ini juga menunjukkan bahwa mengetahui model penetapan harga, mempertimbangkan nilai produk dibandingkan harganya, dan memperjelas biaya pengiriman atau layanan tambahan lainnya, semuanya memengaruhi keputusan konsumen untuk membeli. Untuk menarik perhatian dan mempengaruhi preferensi pelanggan, bisnis harus memiliki kesadaran menyeluruh terhadap elemen ekonomi ini untuk mengubah metode penetapan harga dan promosi mereka di dunia *e-commerce*.

Hasil penelitian (Bahruzen, D, & Nugraha, 2020) mengungkapkan bahwa harga dan kualitas pelayanan mempunyai pengaruh yang paling besar terhadap keinginan konsumen untuk melakukan pembelian lagi. Pemberian diskon dan pelayanan yang cepat berdampak halus pada keinginan pelanggan untuk kembali ke Toko pembelian. Oleh karena itu, harga dan kualitas layanan dapat memberikan keunggulan kompetitif bagi perusahaan ritel di pasar dan menarik pelanggan untuk melakukan pembelian berulang, sesuai dengan implikasi praktis penelitian tersebut. Oleh karena itu, penetapan harga dan penyediaan layanan berkualitas tinggi merupakan faktor penting yang perlu dipertimbangkan untuk mencegah pelanggan beralih ke toko ritel lain yang menyediakan lokasi yang lebih menguntungkan dan pilihan barang yang lebih banyak.

Pelanggan didorong untuk melakukan pembelian berulang di masa depan ketika terdapat kesesuaian antara kinerja produk atau layanan yang diberikan dan kuantitas daya beli konsumen yang dialokasikan untuk suatu produk, yang ditentukan oleh harga. Oleh karena itu, harga dan kualitas layanan dapat memberikan keunggulan kompetitif bagi perusahaan ritel dan menarik pelanggan untuk melakukan lebih banyak pembelian.

2.1.6 *RapidMiner*

Dalam (Nahjan et al., 2023) *RapidMiner* adalah *platform* perangkat lunak yang tangguh untuk pembelajaran mesin dan penelitian data. *RapidMiner* menawarkan serangkaian alat untuk memodelkan, mengevaluasi, menerapkan, dan menyiapkan data. Bahkan mereka yang tidak memiliki keahlian pemrograman dapat membuat dan menguji berbagai model dengan mudah karena arsitektur *RapidMiner* yang ramah pengguna. Dengan antarmuka *drag-and-drop RapidMiner*, pengguna dapat membuat alur kerja untuk pemrosesan dan analisis data. Pekerjaan plot, yang mencakup semua tahapan proses penambangan data, termasuk pembersihan data, pemilihan fitur, dan pemodelan, dibangun menggunakan serangkaian operator yang sudah ada sebelumnya yang juga disertakan dalam paket ini.

Menurut (Faid, 2019) sebuah program yang disebut *Rapidminer* digunakan untuk memproses *data mining*. Penambangan teks *Rapidminer* berfokus pada analisis teks, menemukan pola dalam kumpulan data besar, dan menggabungkannya dengan

database, kecerdasan buatan, dan teknik statistik. Penggunaan operator untuk menghitung data dalam jumlah besar menjadi mudah bagi pengguna dengan *Rapid Miner*. Data dapat diubah dengan operator ini. Untuk menjalankan fungsinya, Rapidminer ditulis dalam bahasa pemrograman Java. *Rapidminer* memiliki sekitar 500 operator. Aspek penting dari analisis data, termasuk pemuatan dan transformasi data, pretreatment dan visualisasi data, pemodelan, dan evaluasi model, tercakup dalam fungsi operator.

Salah satu perangkat lunak yang bersifat terbuka (*open source*) dan tersedia yang digunakan untuk menganalisis pengolahan data adalah *RapidMiner*. Pada *RapidMiner* menggunakan sejumlah metode, termasuk metode deskriptif dan prediktif. *RapidMiner* adalah *open source* dan dapat diakses oleh siapa saja yang menggunakan bahasa pemrograman Java (Sari, Sudewa, Lestari, & Jaya, 2020). Pada penelitian (Sari et al., 2020) *Rapidminer* akan digunakan untuk menghitung teknik *K-Means*. Temuan penelitian tersebut diharapkan dapat membantu pemerintah dalam menentukan derajat kemiskinan di setiap kota atau kabupaten di Provinsi Banten. Oleh karena itu, ketika menerapkan kebijakan untuk mengatasi kesenjangan regional, khususnya terkait kesetaraan pendidikan, pemerintah dapat lebih spesifik dan responsif. Produk akhir dari proses klasterisasi berbantuan *RapidMiner Studio* adalah informasi mengenai kota atau kabupaten mana di Provinsi Tangerang yang memiliki konsentrasi penduduk miskin tertinggi. Informasi ini didukung oleh data tambahan, seperti pengeluaran per kapita dan rata-rata lama pendidikan tiap komunitas.

Dalam (Nofitri & Irawati, 2019) *RapidMiner* merupakan platform perangkat lunak ilmu pengetahuan yang menawarkan lingkungan terpadu untuk penambangan teks (*text mining*), pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), dan analisis prediktif (*predictive analytics*). Program ini mendukung setiap tahapan proses pembelajaran mesin, termasuk persiapan data, visualisasi hasil, validasi, dan pengoptimalan. *RapidMiner* digunakan untuk aplikasi bisnis dan komersial, serta untuk penelitian, pengajaran, pelatihan, pembuatan *prototype* cepat, dan pengembangan aplikasi. Model *open core* digunakan dalam pengembangan *RapidMiner*.

2.2 Penelitian Terdahulu

Tujuan dari kajian ini adalah untuk menemukan kontribusi, metodologi, dan temuan dari penelitian sebelumnya yang dapat digunakan sebagai landasan penelitian ini. Berikut peneliti akan menjelaskan sejumlah metode yang telah digunakan oleh para peneliti sebelumnya, khususnya yang berhubungan dengan penggunaan algoritma Apriori dan K-Means dalam analisis pola pembelian. Selain memberikan konteks teoritis dan praktis untuk permasalahan yang diteliti dalam penelitian ini, penelitian sebelumnya juga digunakan untuk mengidentifikasi kesenjangan penelitian yang perlu diatasi.

Penelitian yang dilakukan (Alinafiah et al., 2018) yang berjudul “Implementasi Data Mining dalam Pengelolaan Stok Obat Menggunakan Metode K-Means

Clustering dan Association Rules Apriori”. Pengelolaan yang tidak efektif menjadi permasalahan dalam penelitian ini, oleh karena itu terdapat kemungkinan terjadinya kelebihan dan kekurangan pasokan. Metode yang digunakan pada penelitian ini adalah *K-Means Clustering* dan *Association Rules Apriori* untuk menampilkan data obat yang sering, umum, dan jarang dibeli serta hubungan antar obat yang dibeli oleh pelanggan. Penelitian ini mendapatkan 3 cluster oleh penggunaan teknik *K-Means Clustering*. *Cluster 1* terdiri dari 231 titik data untuk kelompok yang jarang dibeli, *Cluster 0* terdiri dari 56 titik data untuk kelompok yang dibeli secara rutin, dan *Cluster 2* terdiri dari 41 titik data untuk kelompok yang sering dibeli. Selain itu, metode Apriori yang digunakan dalam studi *Association Rules* mengungkapkan bahwa *METOCLOPRAMID 10mg* dan *METRONIDAZOLE 500mg* adalah obat yang sering dibeli bersamaan. Menurut data, jika seseorang membeli *METOCLOPRAMID 10 mg*, kemungkinan besar dia juga akan membeli *METRONIDAZOLE 500 mg*, dengan tingkat dukungan 17% dan tingkat kepercayaan 62%. Penelitian oleh (Adani et al., 2019) dengan judul “Implementasi *Data Mining* Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma *K-Means Clustering* Pada Toko Syihan” menggunakan metode *Clustering* karena memungkinkan data dapat dikelompokkan berdasarkan atribut yang diinginkan pada Toko Syihan. *K-Means Clustering* adalah algoritma *clustering* yang digunakan dalam aplikasi pemrograman desktop. Dengan penggunaan program *data mining*, hasil dari penelitian ini adalah data penjualan sembako berhasil dikelompokkan menjadi tiga klaster (kelompok), 12 data membentuk

Cluster 1, 10 data membentuk *Cluster* 2, dan 3 data membentuk *Cluster* 3, yang merupakan *cluster* terkecil.

Penelitian yang dilakukan oleh (Supriadi et al., 2021) yang berjudul “Analisis Pola Pembelian Obat di Apotek Menggunakan Algoritma Apriori” bertujuan untuk memberikan informasi kepada apotek tentang bagaimana mempersiapkan persediaan obatnya agar tersedia untuk penjualan yang tinggi agar tidak menumpuk di gudang, meminimalkan kerugian dan menghindari ketidakpercayaan pelanggan dengan menggunakan metode Apriori. Berdasarkan algoritma Apriori pada 350 data transaksi penjualan obat menggunakan alat *RapidMiner* dengan dukungan minimal 5% dan kepercayaan 60% maka dihasilkan 10 aturan asosiasi dengan kombinasi 2 *itemset* dan 3 *itemset*. Penelitian lainnya dilakukan oleh (Satrya Prima Raudini, Lystiowati purwopeni, Dhoni Brilliant & Syahputri, 2021) dengan judul “Penerapan *Data Mining* Terhadap Penjualan Kacamata di Optik Surya Medan Menggunakan Algoritma Apriori”. Tujuan utama dari penelitian ini adalah untuk mengoptimalkan strategi penjualan dan persediaan di Optik Surya Medan berdasarkan transaksi penjualan menggunakan algoritma Apriori. Temuan penelitian ini menunjukkan keefektifan algoritma apriori dalam mengidentifikasi potensi tren pembelian yang dapat dimanfaatkan untuk meningkatkan manajemen rantai pasokan dan taktik pemasaran.

Berdasarkan penelitian oleh (Suseno, 2024) yang berjudul “Analisis Penjualan Produk Menggunakan Algoritma *K-Means* dan Apriori” dijelaskan bahwa penjualan produk di Distro Sextor dianalisis menggunakan *data mining*, dengan teknik *clustering*, dan *association rule* untuk mengidentifikasi tren transaksi penjualan.

Algoritma Apriori digunakan untuk memastikan hubungan antar produk dalam transaksi. Hasil untuk pendekatan *clustering K-Means* didapatkan 5 *cluster*, yaitu: 0, 1, 2, 3, dan 4. Sedangkan untuk analisis menggunakan algoritma apriori dihasilkan 5 *rule* dan 7 produk, yakni SE130, SE111, SE128, SE126, SE04, SE40, dan SE11 yang dapat disarankan sebagai media promosi. Penelitian oleh (Wahyuni, Pane, & Calam, 2023) yang berjudul “Penerapan *Data Mining* dalam Analisa Pola Pembelian pada Penjualan Aksesoris Motor Menggunakan Algoritma Apriori” menggunakan metode *data mining* Algoritma apriori untuk meningkatkan analisis data transaksi dalam mengatasi permasalahan yang disebabkan oleh tren pembelian pelanggan yang tidak menentu dan tidak dapat diprediksi serta penanganan data penjualan yang tidak memadai, yang hanya berfungsi sebagai arsip. Pada penelitian ini dihasilkan aplikasi berbasis web. Omset toko di masa depan dapat ditingkatkan dengan menggunakan temuan perhitungan asosiasi algoritma apriori untuk menentukan taktik penjualan dan persediaan yang paling efektif dan efisien.

Penelitian oleh (Oktaviani, 2024) dengan judul “Implementasi Algoritma Apriori untuk Analisis Pola Pembelian Konsumen pada Toko Serba” menggunakan algoritma apriori yang menggunakan aturan asosiasi untuk mengidentifikasi kombinasi *item* yang memenuhi tingkat dukungan dan kepercayaan yang telah ditentukan. Algoritma apriori berhasil mengidentifikasi pola produk yang dibeli pelanggan secara rutin pada waktu yang sama. Temuan menunjukkan bahwa total 2 item terjual pada waktu yang sama. Di sisi lain, penelitian oleh (Afiasari, Suarna, & Rahaningsi, 2023) yang berjudul

“Implementasi *Data Mining* Transaksi Penjualan Menggunakan Algoritma *Clustering* dengan Metode *K-Means*” menggunakan teknik *data mining* untuk mengkategorikan persediaan produk berdasarkan atribut produk guna mengatasi permasalahan penggunaan informasi persediaan dengan menggunakan metode algoritma *K-Means*. Berdasarkan jumlah 160 *item* produk, kluster 0 menghasilkan 88 *item* yang tergolong laris, *cluster* 1 menghasilkan 26 *item* yang tergolong tidak terjual, dan *cluster* 2 menghasilkan 46 *item* yang tergolong kurang terjual. Tiga *cluster* terbentuk dari hasil teknik *K-Means*.

Penelitian lain dilakukan oleh (Dharshinni, Azmi, Fawwaz, Husein, & Siregar, 2019) dengan judul penelitian “*Analysis of Accuracy K-Means and Apriori Algorithms for Patient Data Clusters*” yang bertujuan untuk melihat pengaruh algoritma *k-means clustering* pada algoritma apriori dengan menggabungkan kedua algoritma. Hasil pengujian menunjukkan kombinasi *algoritma k-means clustering* menghasilkan informasi yang lebih detail dan waktu komputasi yang lebih cepat dibandingkan dengan algoritma apriori total waktu komputasi 21,93 menit dan kombinasi algoritma *k-means clustering* 17,41 menit. Dalam penelitian (Sinaga & Yang, 2020) yang berjudul “*Unsupervised K-Means Clustering Algorithm*” peneliti mengusulkan skema baru dengan pembelajaran kerangka kerja untuk algoritma pengelompokan *k-means*. Peneliti mengadopsi manfaat ketentuan tipe entropi *k-means* untuk membangun skema persaingan. Algoritma *K-Means* yang diusulkan menggunakan jumlah poin sebagai jumlah awal *cluster* untuk penyelesaian masalah inisialisasi. Penelitian ini disimpulkan

bahwa meskipun *K-Means* adalah algoritma yang lugas dan efektif, algoritma ini dapat dibuat lebih akurat dan serbaguna dengan penyesuaian lebih lanjut, terutama bila digunakan pada data dengan struktur yang rumit.

Menurut penelitian (Setiawan et al., 2020) dengan judul “Pemanfaatan Metode *Association Rules* dan *Holt-Winter Multiplicative* untuk Meningkatkan Peluang Penjualan Obat Pertanian pada PT. ABC” peneliti menggabungkan pendekatan *Holt-Winter Multiplicative* untuk mengevaluasi perubahan tren pada data musiman dengan metode *Association Rules* untuk mengkaji tren dan mencari setiap kemungkinan untuk meningkatkan penjualan. Dengan menggabungkan metode *Association Rules* dan *Holt-Winter, Multiplicative* menghasilkan data mengenai cara meningkatkan potensi penjualan dengan meramalkan perubahan stok dan menyesuaikan dengan kebutuhan pelanggan serta prediksi kondisi area pemasaran. Dalam (Suryanto & Adevi, 2023) dengan judul penelitian “Analisa Penjualan Toko Hijab Kiki Hn dengan Penerapan *Data Mining* Metode *K-Means Clustering* dan *Market Basket Analysis*” melakukan penelitian untuk mengidentifikasi barang-barang yang laku, mana yang tidak, dan untuk menemukan jawaban bagaimana bisnis dapat menghindari kesulitan dalam mengklasifikasikan produk yang dibutuhkan pelanggan dan penyimpanan data yang kurang efisien dengan menggunakan metode *Clustering* dan algoritma Apriori. Berdasarkan temuan penelitian, terdapat 5 produk dalam kategori sangat laris, 4 produk laris, dan 6 produk dalam kategori kurang laris. Salah satu pola asosiasi dari 30 hasil pengolahan data transaksi menggunakan *Market Basket Analysis* memenuhi kriteria,

khususnya jika membeli Pashmina maka akan membeli Bros dengan tingkat akurasi tertinggi sebesar 72,22%, dan desain yang diperoleh dapat dimanfaatkan untuk memberikan diskon atau pembelian bundel.

Penelitian oleh (Andrianto & Elisa, 2021) tentang “*Analysis Clustering* dengan *K-Means* untuk Pengelompokan Penjualan Elektronik pada Batam IT Mart” bertujuan untuk membantu Batam IT Mart dalam mengklasifikasikan penjualan dan mengidentifikasi informasi tentang jenis barang yang diminati dan tidak diminati. *Metode K-means Clustering* yang digunakan dalam penelitian ini adalah dengan membagi data menjadi beberapa kelompok kemudian melakukan proses komputasi sehingga menghasilkan temuan berupa kelompok data tentang jenis-jenis barang yang laris, tidak laris, dan sangat laris. Hasil penelitian yang didapat adalah terdapat 3 *cluster*, *Cluster 0* berjumlah 19 *item*, *Cluster 1* berisi 2 *item*, dan *Cluster 2* berisi 4 *item*. Oleh karena itu, analisis dan pengelompokan data penjualan menggunakan pendekatan *k-means clustering* dapat membantu proses penelitian ini. Penelitian lainnya adalah oleh (Sigalingging & Harman, 2020) dengan judul “Analisis Pola Pembelian Produk Menggunakan Algoritma Apriori” dengan tujuan untuk mengetahui tren pembelian konsumen dan mengetahui bagaimana keadaan di minimarket. Untuk mengelola data yang diperoleh dicari pola keterkaitan antar objek yang biasa terjadi dalam beberapa transaksi dengan menggunakan metode aturan asosiasi. *Data mining* digunakan bersama dengan metode algoritma apriori untuk memastikan tren pembelian produk. Hasil penelitian menunjukkan bahwa nilai *support* sebesar 46,66% dan nilai keyakinan

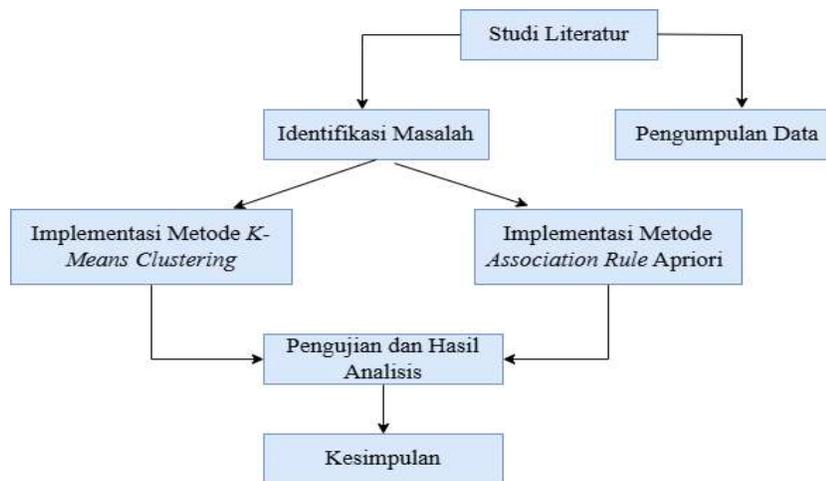
sebesar 87,50 untuk barang telur dan indomie, serta nilai dukungan sebesar 46,66% dan nilai keyakinan sebesar 63,63% untuk barang indomie dan telur.

Penelitian yang dilakukan oleh (Novianti & Elisa, 2020) yang berjudul “Penentuan Aturan Asosiasi Pola Pembelian pada Minimarket dengan Algoritma Apriori” menemukan pola asosiasi berdasarkan perilaku belanja pelanggan sehingga peneliti dapat mengidentifikasi produk yang dibeli secara bersamaan dengan menggunakan analisis aturan algoritma apriori. Temuan penelitian ini adalah dengan nilai dukungan 40% dan tingkat kepercayaan 85,71%, Telur dan Mie memiliki nilai dukungan dan kepercayaan tertinggi.

Berdasarkan penjelasan penelitian terdahulu, penelitian ini diharapkan dapat menghasilkan metode yang lebih tepat dan ideal dalam mengenali tren pembelian konsumen yang akan membantu Toko Elektronik A8 dan Toko Furnitur Piayu mengambil keputusan bisnis yang lebih baik.

2.3 Kerangka Pemikiran

Kerangka pemikiran adalah peta jalan yang menguraikan pendekatan atau prosedur metodologis dalam melakukan penelitian agar analisis dan pemecahan masalah dapat dilakukan secara jelas, konsisten, dan selaras dengan tujuan. Berikut adalah kerangka pemikiran yang dibuat peneliti dalam penelitian ini:



Gambar 2. 2 Kerangka Pemikiran

Sumber: Peneliti, 2024

Uraian dari kerangka penelitian diatas adalah sebagai berikut:

1. Studi Literatur, memahami latar belakang penelitian, mengidentifikasi kesenjangan pada penelitian sebelumnya, dan mengembangkan landasan teori yang mendukung penelitian yang direncanakan merupakan tujuan studi literatur.
2. Identifikasi Masalah, tahapan ini peneliti mengidentifikasi, menemukan, dan mendefinisikan permasalahan utama yang menjadi pokok kajian. Tahap ini berupaya untuk memahami sepenuhnya kesulitan atau kesenjangan yang ada dalam konteks atau keadaan tertentu. Menemukan masalah membantu peneliti dalam menentukan fokus dan arah penelitian untuk memastikan bahwa hasilnya relevan dan tepat.

3. Pengumpulan Data, pada tahap ini peneliti mengumpulkan data transaksi konsumen dari Toko A8 Electronic & Furniture Piayu, seperti detail produk, harga pembelian, dan waktu transaksi.
4. Implementasi Metode *Association Rule* Apriori, Setelah konsumen dibagi menjadi beberapa *cluster*, data transaksi di setiap *cluster* diterapkan pada algoritma Apriori, yang mengidentifikasi korelasi antara produk yang sering dibeli dalam setiap kelompok pelanggan.
5. Implementasi Metode *K-Means Clustering*, metode *K-Means* digunakan untuk mengolah data pelanggan guna membagi pelanggan menjadi beberapa kelompok (*cluster*) sesuai dengan atribut pembeliannya, seperti jumlah dan frekuensi transaksi.
6. Pengujian dan Hasil Analisis, untuk mempelajari lebih lanjut tentang kebutuhan dan preferensi setiap segmen konsumen, pola pembelian yang ditemukan dianalisis.
7. Kesimpulan, Berdasarkan hasil analisis, didapatkan kesimpulan serta saran diberikan untuk membantu pengecer dalam mengelola produk dan strategi pemasaran mereka dengan lebih sukses. Beberapa saran tersebut mencakup penempatan produk yang lebih bertarget, menggabungkan promosi, dan menyesuaikan penawaran untuk setiap pelanggan *cluster*.

2.4 Hipotesis Penelitian

Berdasarkan hasil kerangka pemikiran di atas, hipotesis penelitian ini adalah:

1. Diduga dengan menggunakan algoritma Apriori, beberapa barang yang dibeli dalam satu transaksi, pasangan produk tersebut akan ditemukan, dan dapat dimanfaatkan untuk meningkatkan penjualan melalui promosi atau strategi penempatan produk.
2. Diduga algoritma *K-Means* dapat menunjukkan bahwa terdapat segmen pelanggan yang dengan atribut belanja yang berbeda-beda, termasuk frekuensi pembelian, jenis produk, dan jumlah pembelanjaan.