

BAB II

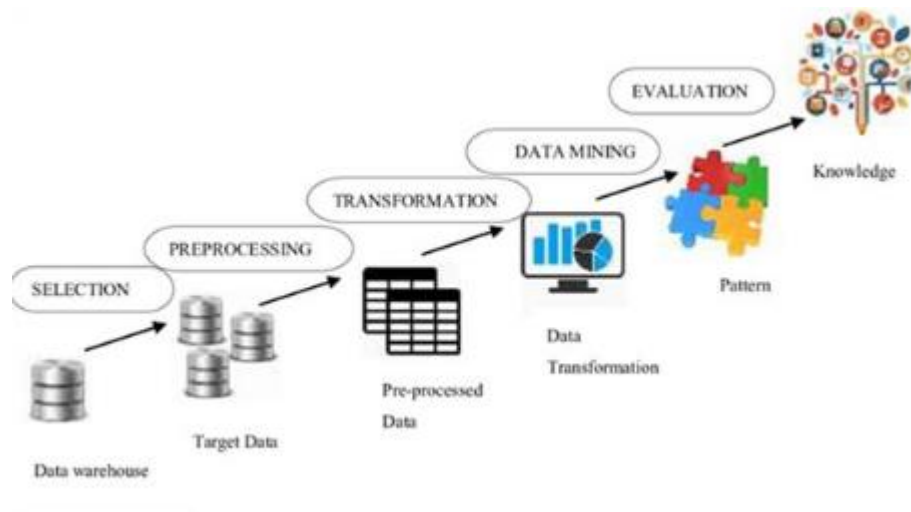
TINJAUAN PUSTAKA

2.1 Teori Dasar

2.1.1 KDD (Knowledge Discovery in Databases)

KDD merupakan suatu proses yang sistematis dan menyeluruh untuk mendapatkan pengetahuan yang bermanfaat, tersembunyi, dan dapat diinterpretasikan dari kumpulan data yang besar (Pebdika, Herdiana, and Solihudin 2023). *Knowledge Discovery in Database* (KDD) adalah istilah yang sering digunakan untuk menggambarkan proses penggalian informasi tersembunyi dalam database besar (Viska and Elisa 2023). Proses *Knowledge Discovery in Database* (KDD) yang dapat dilakukan ulang terus-menerus untuk mendapatkan hasil terbaik (Sitanggang et al. 2024). KDD tidak hanya menemukan pola atau hubungan dalam data, tetapi juga memastikan bahwa informasi yang dihasilkan berguna untuk pengambilan Keputusan (Apriyani, Dikananda, and Ali 2023).

Untuk lebih jelasnya tahapan KDD bisa dilihat dari gambar berikut (Wartumi, Kurniawan, and Wijaya 2024) :



Gambar 2. 1 tahapan dalam KDD

1. Selection (pemilihan data)

Pemilihan data, atau pemilihan data, adalah tahap pertama proses KDD. Pada tahap ini, data yang dianggap relevan dan sesuai dengan tujuan analisis dipilih dari kumpulan data yang tersedia.

2. Preprocessing data

Preprocessing atau pra-pemrosesan data dilakukan setelah data yang relevan telah dipilih. Tahap ini bertujuan untuk mengatasi masalah umum seperti data hilang (nilai yang tidak ada), duplikat (data ganda), nilai ekstrem (outliers), dan inkonsistensi format atau isi.

3. *Transformation* (transformasi)

Transformasi atau transformasi data adalah langkah ketiga dalam proses KDD. Data yang telah dibersihkan akan diubah ke dalam struktur atau format yang lebih cocok untuk proses analisis berikutnya

4. Data mining

Pada titik ini, diharapkan pola yang ditemukan akan memberikan nilai tambahan dalam bentuk wawasan strategis yang akan membantu pengambilan keputusan.

5. Interpretation / evaluasi data

Interpretation and evaluation adalah tahapan terakhir dari proses KDD. Ini adalah proses untuk menilai pola yang telah ditemukan dan menilai seberapa valid, berguna, dan mudah dipahami dalam konteks aplikasinya.

2.1.2 Data Mining

Data mining adalah proses analitis yang digunakan untuk mengekstraksi pengetahuan yang berharga, konsisten, dan tersembunyi dari kumpulan data yang besar (Nugroho, Ma'arif, and Arif 2022). *Data mining* adalah proses mengekstraksi dan mengidentifikasi informasi penting dan relevan dari berbagai database yang terkait dengan menggunakan metode matematika, statistik, kecerdasan buatan, dan pembelajaran mesin (Surbakti 2021). *Data mining* digunakan untuk menemukan pola yang tepat untuk mengekstrak data berharga dari database yang sangat besar (Almufqi and Voutama 2023). Tahap perolehan data, atau proses penambang data, adalah data mining. Tujuannya adalah untuk menemukan pola-pola yang menarik dan mendapatkan pengetahuan dari suatu dataset yang memiliki banyak data (Rahmawati and Nopriadi 2020).

2.1.3 Association Rules (Aturan Asosiasi)

Aturan asosiasi, juga dikenal sebagai aturan asosiasi, adalah salah satu teknik data mining yang digunakan untuk menemukan hubungan atau keterkaitan antar item dalam kumpulan data (Rasid Siddik 2024). Teknik ini sangat populer dalam

analisis perilaku konsumen dan analisis pasar keranjang, di mana tujuannya adalah untuk menemukan kombinasi produk yang sering dibeli oleh pelanggan secara bersamaan. Menurut (nindy sari 2022), tujuan penambangan aturan asosiasi adalah untuk menemukan aturan implikasi dari bentuk $X \Rightarrow Y$, di mana X dan Y adalah himpunan item, dan aturan tersebut menunjukkan bahwa ada kemungkinan besar bahwa jika X muncul dalam sebuah transaksi, Y juga akan muncul. Association rules adalah Metode penggabungan aturan memasukkan kumpulan data besar, atau kumpulan item, ke dalam kumpulan aturan. Untuk memastikan bahwa aturan asosiasi dapat digunakan, perlu dievaluasi dukungan dan kepercayaan dari aturan yang dibentuk menurut (Amalia and Setiawansyah 2021) Metode Peraturan Asosiasi menggunakan algoritma apriori untuk mengidentifikasi dan memproses data transaksi penjualan untuk menghasilkan rekomendasi penjualan yang paling banyak untuk mempersiapkan strategi pemasaran bulan berikutnya (Lienata et al. 2021).

Menurut (Khirdany et al. 2022) Association rules memiliki karakteristik sebagai berikut:

1. Peraturan Asosiasi tidak bersifat prediktif (seperti klasifikasi), tetapi deskriptif
2. Bisa membuat banyak aturan, jadi harus memilih yang paling relevan.
3. Tidak membutuhkan label atau kelas (pendidikan tidak diawasi)
4. Berhasil melakukan analisis pola berulang pada kumpulan data yang besar.

2.1.4 Algoritma Apriori

Salah satu algoritma yang digunakan untuk menemukan pola frekuensi tinggi yang sangat terkenal adalah algoritma apriori (Albab and Hidayatullah 2022). Algoritma Apriori bertujuan untuk menemukan set barang sering dengan menggunakan sekumpulan data mining untuk mempelajari bagaimana pelanggan membeli barang secara bersamaan (Ade Irma Amanda, Debi Setiawan, and Liza Trisnawati 2023). Pola item dalam database yang memiliki frekuensi atau dukungan di atas ambang batas tertentu disebut pola frekuensi tinggi yang disebut sebagai minimum dukungan. Prinsip apriori, yang menyatakan bahwa jika suatu set item tidak memenuhi ambang minimum dukungan, maka semua superset dari set tersebut juga tidak akan memenuhi ambang tersebut (Febiana and Alda 2024).

Salah satu teknik pengambilan data yang efektif adalah metode apriori, yang terdiri dari aturan asosiasi yang menggabungkan itemset penjualan merek dengan nilai pendukung yang lebih besar dari minimal nilai pendukung dari batas nilai kepercayaan (Handoko et al. 2024). Algoritma Apriori berguna untuk analisis pembelian dan persediaan karena membantu mengelola data dengan menemukan pola hubungan antara item melalui penghitungan dukungan dan keyakinan (Parinduri, Defit, and Nurcahyo 2024). Selain itu, algoritma apriori dianggap sebagai solusi yang menguntungkan untuk pemecahan masalah (Merliani et al. 2022). Algoritma Apriori adalah sebuah algoritma data mining metode asosiasi, digunakan untuk mengidentifikasi pola item produk yang sering dibeli oleh pelanggan pada saat yang sama atau pola dalam data (nindy sari 2022).

Menurut (Musdalifah and Jananto 2022) rumus yang akan digunakan dalam algoritma apriori adalah sebagai berikut :

Support :

$$\text{Support } A = \frac{\text{Jumlah transaksi mengandung } A}{\text{Total transaksi}} \times 100\%$$

Rumus 2.1 Nilai *Support*

Kemudian, nilai *support* 2 item diperoleh menggunakan rumus:

$$\text{Support } A, B = \frac{\text{Total transaksi mengandung } A \text{ dan } B}{\text{Total transaksi}} \times 100\%$$

Rumus 2.2 Nilai *Support* 2 item

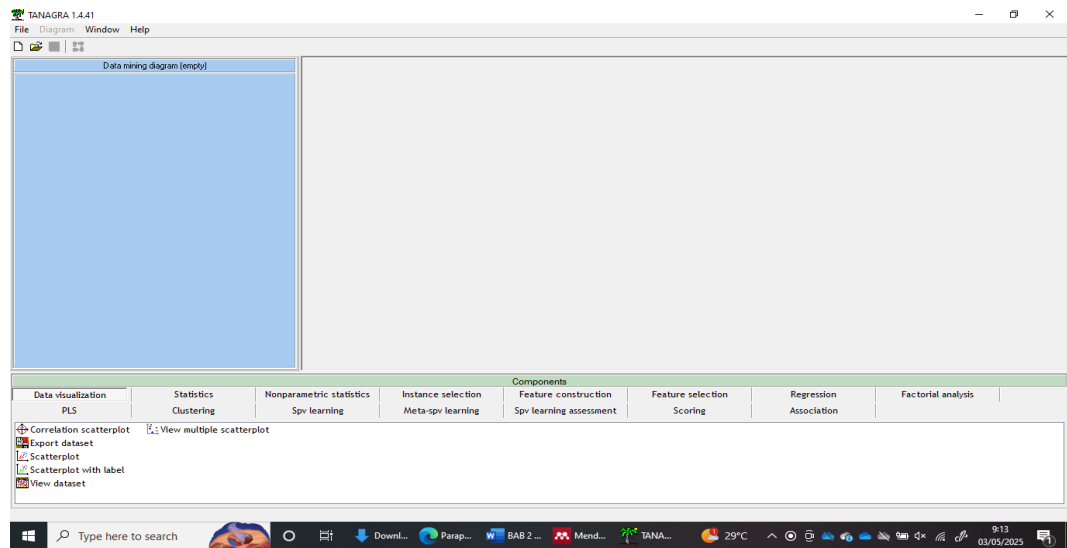
Confidance:

$$\text{Confidance } P(B|A) = \frac{\text{Total transaksi mengandung } A \text{ dan } B}{\text{Total transaksi mengandung } A} \times 100\%$$

Rumus 2.3 Nilai *Confidance*

2.1.5 Tanagra

Tanagra adalah perangkat lunak open-source yang dirancang untuk membantu proses data mining dan eksplorasi. Dengan antarmuka grafisnya, (Primadona and Fauzi 2023). Tanagra populer di kalangan akademisi dan praktisi yang membutuhkan alat yang sederhana namun kuat untuk analisis data dalam bisnis, kesehatan, ilmu sosial, dan banyak bidang lainnya (Apriandi, Sari, and Sarif 2024) berikut adalah tampilan awal dari Tanagra.



Gambar 2. 2 tampilan tanagra

2.1.6 Usaha Retail

Usaha retail adalah keseluruhan aktivitas jual beli barang atau jasa secara langsung kepada konsumen yang membutuhkan barang atau jasa untuk memenuhi kebutuhan hidupnya (Hofipah et al. 2023). Bisnis retail dapat menjual barang mereka dengan berbagai cara, seperti di toko (grosir,minimarket),di internet, secara langsung, atau melalui layanan pos dan telepon (Etika et al. 2024). Dalam menjalankan usaha retail apalagi usaha retail dalam bentuk toko diperlukan system pengelolaan yang baik,agar semua penempatan barang tersusun rapi sesuai stok.Pengelolaan adalah sebagai proses khusus yang terdiri dari tindakan perencanaan, pengorganisasian, pergerakan, dan pengendalian yang dilakukan untuk menentukan dan mencapai tujuan yang telah ditentukan dengan menggunakan sumber daya manusia dan sumber daya lainnya (Rinta, Febriana, and Wulandari 2022). Tujuan pengelolaan adalah untuk memastikan bahwa biaya

penyimpanan barang, biaya pemesanan, dan tingkat layanan yang diberikan kepada pelanggan seimbang (Manik and Marbun 2021).

2.2 Penelitian Terdahulu

Penelitian sebelumnya memberikan konteks dan pijakan untuk penelitian ini. Peneliti dapat memperoleh pemahaman yang lebih baik tentang masalah yang dibahas, metode yang digunakan, dan hasil yang dicapai dengan melihat penelitian sebelumnya yang membahas topik sejenis. Selain itu, penelitian ini memungkinkan peneliti untuk menemukan kekurangan atau kelemahan dalam penelitian sebelumnya, yang memungkinkan mereka untuk membuat kontribusi baru yang relevan dan unik.

Penelitian yang dilakukan oleh (Tarigan et al. 2022) Toko sinar harahap mengalami kerugian karena stok barang tidak laku dan habis terjual, terutama barang yang sudah kadaluarsa dan mudah rusak. Sebaliknya, jika stok barang lebih sedikit daripada permintaan pelanggan, pelanggan akan meninggalkan toko karena barang yang akan mereka beli tidak mencukupi, item yang dijual ditoko ini berupa bahan sembako. peneliti pada penelitian ini menggunakan algoritma apriori untuk pemecahan permasalahan, dari hasil perhitungan didapatkan hasil Dengan nilai kepercayaan sebesar 62%, pembelian beras dan telur mungkin memiliki permintaan tertinggi. Selanjutnya, penelitian yang dilakukan oleh (Dongga et al. 2023) Toko Swapen Jaya mungkin mengalami kerugian karena stoknya lebih banyak daripada yang diinginkan pelanggan.. Ini adalah situasi yang sering terjadi ketika pelanggan ingin membeli barang, tetapi barang yang mereka inginkan ternyata tidak ada di gudang. Hasil penelitian ini sembako dan minuman memiliki nilai dukungan 5%

dan nilai dukungan 36%, sedangkan hasil makanan dan minuman memiliki nilai dukungan 11% dan nilai dukungan 22%.

Penelitian selanjutnya oleh (Sari et al. 2024) Karena tidak familiar dengan tugas dan jumlah barang yang banyak menjadi referensi, operator komitmen masih sering menghadapi masalah dalam menentukan kodefikasi barang. Akibatnya, penelitian ini akan mencari solusi atau menggunakan algoritma untuk menentukan kodefikasi barang yang tepat. Hasil yang didapatkan adalah Menurut hasil perhitungan, algoritma Apriori dan FP-Growth dapat menghasilkan 85 aturan asosiasi dengan nilai dukungan di atas 50% dan kepercayaan 90%. Penerapan metode ini dalam Aplikasi SAKTI akan membantu operator dalam menentukan kodefikasi barang yang tepat. Selanjutnya penelitian oleh (Nurhidayanti and Kurniawati 2022) dengan judul Sebagai showroom dan bengkel, Honda Tanabang Motor terus melakukan pencatatan barang keluar-masuk dan pengecekan stok secara manual. Akibatnya, data yang disajikan menjadi tidak valid. Akibatnya, pemesanan stok barang tidak efektif yang mengakibatkan banyak barang yang sudah kosong tetapi tidak dipesankan. Hasil yang diperoleh dari penelitian ini adalah tiga aturan yang memiliki nilai kepercayaan tertinggi sebesar 75% .

Penelitian yang dilakukan oleh (Ginting, Saripurna, and Fitriani 2021) Perusahaan penjualan barang lainnya sering menghadapi masalah oprasional dengan inventaris, atau persediaan barang Hasil yang didaptakan adalah Jika pelanggan membeli Teh Botol, mereka akan menerima dukungan air 16% dan keyakinan 62%, dan perusahaan/gudang harus menyimpan barang tersebut. Jika pelanggan membeli merek Bear, mereka akan menerima dukungan air 16% dan

keyakinan 60%, dan perusahaan/gudang harus menyimpan barang tersebut.,Jika pelanggan membeli Chitato, mereka akan menerima dukungan Cheetos 18% dan keyakinan 100%, dan perusahaan/gudang harus menyimpan barang tersebut.

Penelitian oleh (Panjaitan 2021) Salah satu masalah yang sering terjadi di UD.Chandra Jaya Tani adalah kehabisan stok barang karena tingkat belanja konsumen yang tidak beraturan dan penumpukan barang.Berdasarkan hasil perhitungan didapatkan Konsumen yang membeli produk Ally akan membeli Bablas dengan nilai dukungan 30,61% dan keyakinan 65,21%. Mereka juga akan membeli Basmilang dengan nilai dukungan 14,28% dan keyakinan 39,13%. Mereka juga akan membeli Basf dengan nilai dukungan 10,20% dan keyakinan 21,73%.

Selanjutnya, penelitian oleh (Saraswati, Fauziah, and Nathasia 2023) Sebuah sistem untuk mencatat barang masuk dan keluar di Kementerian Pariwisata dan Ekonomi Kreatif menampilkan stok 0 dan minus. Selanjutnya penelitian yang dilakukan oleh (Syahara, Adiha, and Windarto 2021) proses menggunakan algoritma apriori *Asociation Rule*. Dengan menggunakan aplikasi pendukung Tanagra 1.4.50, pola kombinasi sistem persediaan Pasir dan Semen dari Toko Bahan Bangunan Karang Sari memiliki nilai kepercayaan sebesar 75%.

Penelitian yang dilakukan oleh (Daeli, Rahayu, and Hadinata 2023) Toko Santi Fotokopi masih menggunakan catatan manual untuk memprediksi barang yang akan dimasukkan kembali ke dalam stok. Hasil analisis yang dilakukan dengan menggunakan perbandingan minimum dukungan sebesar 10% dan minimum keyakinan sebesar 60%, menghasilkan enam aturan yang terkait dengan

pola penjualan tertentu yang memiliki nilai keyakinan tertinggi, yaitu penjualan kertas Binder, yang memiliki nilai keyakinan sebesar 70%. Dengan menggunakan enam aturan ini, pemilik Santi Fotokopi dapat menyediakan stok produk tersebut, yang akan menghasilkan keuntungan bagi toko. Penelitian selanjutnya dilakukan oleh (Alamsyah 2024) Dibutuhkan sistem yang dapat membantu restoran mengetahui stok bahan baku apa saja yang harus disiapkan dan diletakkan secara bersamaan. Hasil yang diperoleh adalah Hasil percobaan menunjukkan bahwa bahan baku gembung akan menghasilkan 80% ikan nila, 83% ikan krapu, dan 100% ikan krapu jika bahan baku gembung digunakan.

Penelitian oleh (Sibarani 2020) Toko akan mengalami kerugian jika stok barangnya terlalu banyak dibandingkan dengan permintaan pelanggannya karena banyaknya barang yang tidak terjual, terutama barang yang sudah kadaluarsa atau mudah rusak. Untuk hasil dari permasalahan ini Berdasarkan temuan penelitian ini, dapat disimpulkan bahwa algoritma Apriori terbukti efektif dalam penambangan data karena dapat memprediksi hasil penjualan bahan makanan dan memberi tahu Anda tentang barang apa yang seharusnya ada di toko. Algoritma ini berhasil menemukan pola pembelian yang umum, seperti permintaan yang tinggi untuk telur, minyak, dan tepung. Ini ditunjukkan oleh nilai keyakinan sebesar 96%. Selanjutnya penelitian yang dilakukan oleh (Ardanu 2022).

Penelitian yang dilakukan oleh (Alamsyah 2024) Metode algoritma apriori, yang digunakan untuk membuat keputusan tentang pemenuhan persediaan barang, adalah salah satu contohnya. Hasil yang didapatkan dari perhitungan apriori ini adalah Dalam kasus ini, data penjualan Arfishop diuji dengan 21 aturan asosiasi dukungan

dan kepercayaan. Salah satu contohnya adalah aturan "Zippo blue ice", yang mengatakan bahwa minyak zippo memiliki nilai kepercayaan 100%. Toko Arfishop dapat menggunakan hasil asosiasi algoritma apriori ini untuk melakukan prediksi stok persediaan barang dengan efisien.

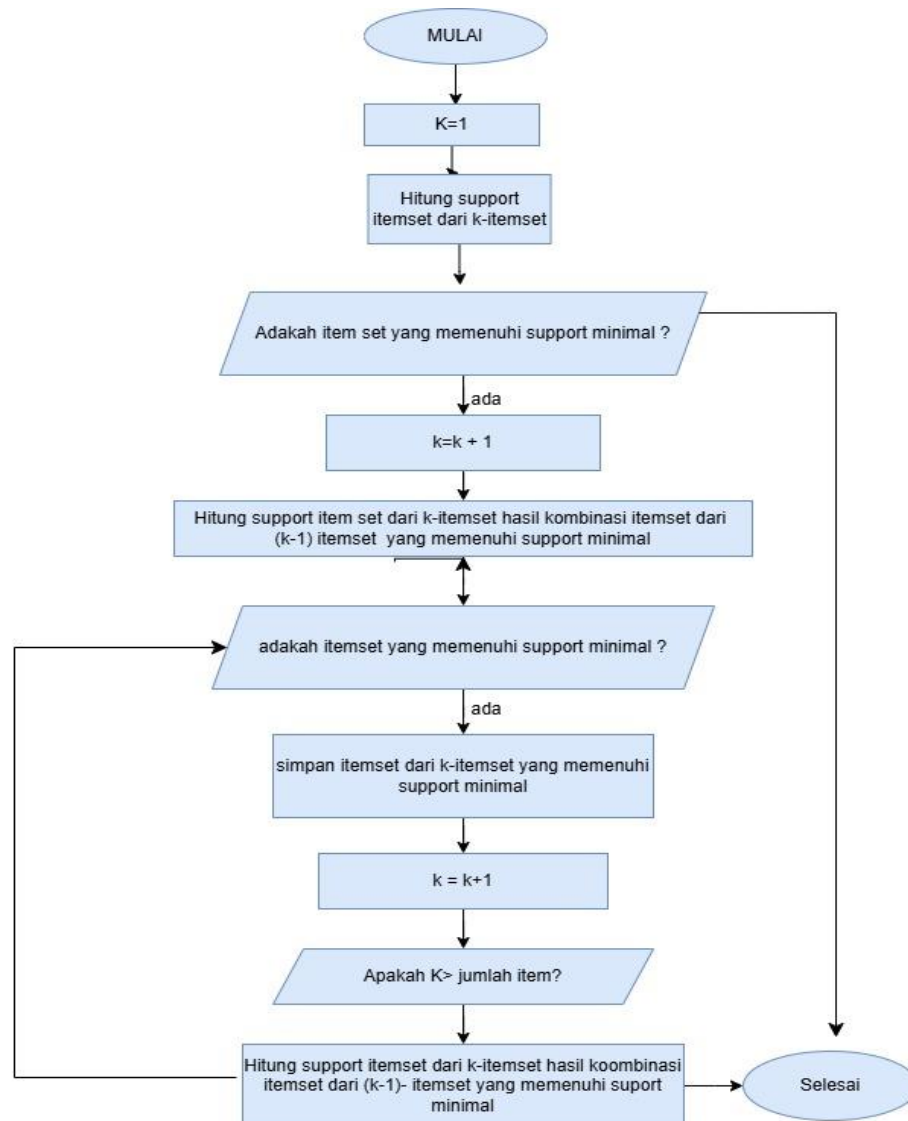
Selanjutnya penelitian oleh (Rika Widianita 2023) Tidak adanya pengolahan data di toko Bumkammart dapat membahayakan keuntungan toko dan pembeli. Pembeli akan menganggap kecewa ketika barang yang diinginkannya kehabisan, biasanya karena banyak peminat dan stok terbatas. Oleh karena itu digunakan metode algoritma apriori yang menghasilkan hasil Menurut literasi satu item set, barang yang paling sering dibeli adalah ATK dengan nilai dukungan 83,01%, minuman dengan nilai dukungan 90,20%, perlengkapan RT dengan nilai dukungan 80,39%, dan snack dengan nilai dukungan 84,97%. Barang-barang yang sering dibeli ditemukan dalam kombinasi dua kategori: ATK-Minuman dengan nilai dukungan 76,47%, ATK-Perlengkapan RT dengan nilai dukungan 70,59, ATK-Snack dengan nilai dukungan 72,55%, Minuman-Perlengkapan RT dengan nilai dukungan 74,51%, Minuman-Snack dengan nilai dukungan 80,39%, dan Perlengkapan RT-Snack dengan nilai dukungan 71,90%.

Penelitian yang dilakukan oleh (Riszky and Sadikin 2019) Dalam strategi pemasaran produknya, berdasarkan perhitungan peneliti mendapatkan hasil Pemasaran dapat menggunakan strategi seperti bundling produk (di mana grapeola, jeruk, dan pear dijual dalam satu paket hemat), buy 2 get 1 (di mana jika Anda membeli 2 kg jeruk, Anda akan mendapatkan 1 kg pear), dan diskon produk (jika Anda membeli lebih dari 5 kg jeruk, Anda akan mendapatkan diskon 5%). Hasil

penelitian ini dapat digunakan oleh organisasi untuk mendukung strategi pemasaran merek.

2.3 Kerangka Pemikiran

Berikut adalah kerangka pemikiran dari penelitian ini:



Gambar 2. 3 Kerangka Pemikiran

2.4 Hipotesis Penelitian

Dari kerangka pemikiran diatas, adapun hipotesis yang ada pada penelitian ini adalah :

1. Penerapan algoritma Apriori pada data transaksi penjualan Toko Grosir Abang Putra akan berhasil mengidentifikasi pola hubungan asosiasi yang signifikan antar barang yang sering dibeli bersamaan.
2. Kelompok barang (*itemset*) yang memiliki frekuensi kemunculan tinggi dalam transaksi penjualan Toko Grosir Abang Putra, sebagaimana diidentifikasi oleh algoritma Apriori, dapat memberikan informasi berharga bagi strategi penataan produk dan promosi di toko tersebut.