

BAB II

LANDASAN TEORI

2.1. Tinjauan Pustaka

Tinjauan pustaka dalam penelitian ini adalah artikel ilmiah yang berjudul Pencarian Pola Penjualan Minuman Kemasan di toko Renji Semarang Menggunakan Algoritma Apriori, Andriyani Jananto, (2020), penelitian pada toko Renji Semarang memfokuskan pada penyelesaian masalah di toko Renji yakni si pemilik toko kesulitan mengolah data. transaksi penjualan di toko tersebut dan mengalami kesulitan dalam mengolah ketersediaan barang, oleh karena itu penjualan tidak maksimal. Maka dari itu pihak Toko Renji membutuhkan informasi pola kombinasi produk minuman kemasan agar dapat memenuhi persediaan produk minuman yang saling berkaitan. Dengan adanya implementasi data mining menggunakan Algoritma Apriori pada Toko Renji dapat membantu pemilik toko dalam mengetahui pola kombinasi produk minuman yang saling berhubungan. Menganalisa transaksi penjualan tersebut dapat menggunakan teknik asosiasi tujuannya agar bisa melihat pola hubungan antar barang yang dibeli konsumen secara bersamaan.

Software yang akan digunakan dalam menentukan aturan asoasiannya yakni menggunakan RStudio pada tahapan ini dilakukan 6 tahapan penelitian ini yaitu: 1) Import data, 2) Cleaning data, 3)Melihat Jumlah Transaksi, 4)Membuat Pola Frekuensi, 5)Pembentukan Rules Apriori, 6) Pemanggilan

Berdasarkan hasil penelitian yang telah dilaksanakan dapat ditarik kesimpulan bahwa Implementasi yang telah dilakukan dengan menggunakan Software Rstudio dengan nilai support 1% dan nilai confidence 50% menghasilkan 14 aturan (rules) asosiasi dalam transaksi penjualan minuman kemasan di Toko Renji. Nilai confidence tertinggi yaitu 100% pada pembelian

Kopi Cup dan Vit 1500ml maka akan membeli Le Minerale 600ml secara bersamaan.

Tinjauan pustaka dalam penelitian ini adalah artikel ilmiah

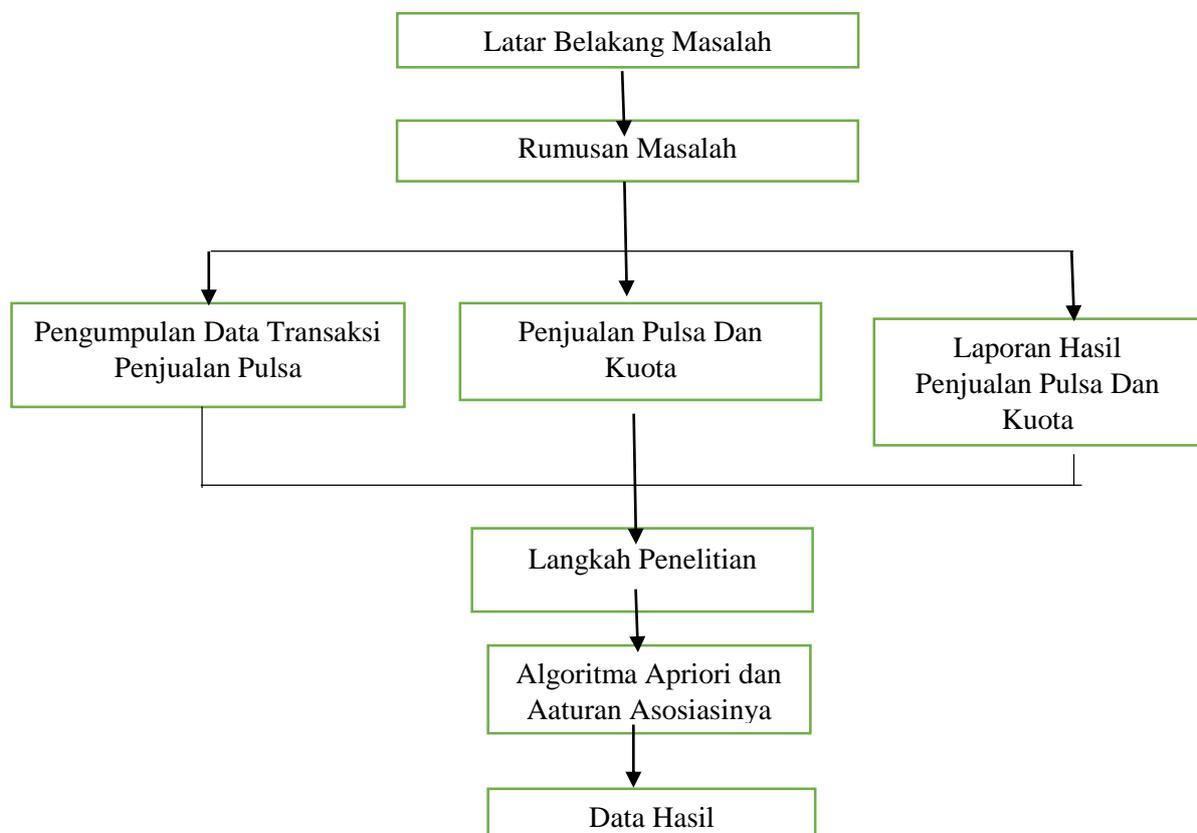
yang berjudul “Metode Data Mining Terhadap Data Penjualan Sparepart Mesin Fotocopy Menggunakan Algoritma Apriori”. *Rahmawati Merlina, (2018)* Penelitian ini dilakukan untuk mengetahui produk sparepart mesin fotocopy dengan penjualan terbanyak yang terjual secara bersamaan dibutuhkan metode agar didapatkan informasi baru. Yaitu dengan menggunakan algoritma Apriori dan bantuan tools Tanagra. Algoritma Apriori adalah jenis aturan asosiasi pada data mining. Dengan menggunakan perhitungan algoritma Apriori maka dapat ditentukan produk sparepart mesin fotocopy yang banyak terjual secara bersamaan yang bermanfaat dalam penyelesaian strategi pemasaran agar lebih meningkatkan tingkat penjualan. Dari hasil penelitian ini dapat ditemukan produk yang paling banyak terjual yaitu drum dc dan toner, dengan menghasilkan kesimpulan bahwa produk Toner dan Drum dc dengan nilai support 50% dan nilai confidence 66,67% atau produk Drum Dc dan Toner dengan nilai support 50% dan nilai confidence 85,71%. Dalam hal ini peneliti menggunakan beberapa tahapan yaitu: 1) Pembentukan itemset, 2) Daftar Support dari tiap 1 itemset, 3)Kombinasi 2 itemset, 4)Daftar Support Dari 2itemset, 5)Kombinasi 3itemset, 6)Pembentukan Asosiasi Final, 7)Implementasi menggunakan Software Tenagra.

Tinjauan pustaka dalam penelitian ini adalah artikel ilmiah yang berjudul Penerapan Algoritma Apriori Pada Data Penjualan Barbar Warehouse *Lestari Hafiz, (2020)*

Dunia bisnis para pelaku usaha saling bersaing dalam berinovasi dan melakukan strategi untuk menarik pelanggan guna meningkatkan penjualan. Barbar Warehouse merupakan sebuah retail online yang menjual produk fashion, kebutuhan bayi seperti popok, seprei untuk kebutuhan para pelanggan. Barbar Warehouse sendiri mempunyai 5 toko yang terdaftar dalam

marketplace Zilingo. Strategi Barbar Warehouse sebagai retail online melakukan pembelian dalam 3 bulan sekali, dalam waktu tersebut Barbar. Data transaksi penjualan setiap harinya selalu bertambah, semakin banyak data transaksi tersimpan menyebabkan penyimpanan data menjadi besar. Data transaksi penjualan dimanfaatkan dan diolah menjadi informasi dalam meningkatkan penjualan produk. Barbar Warehouse memerlukan sebuah metode untuk menganalisa pangsa pasar melalui pola penjualan untuk mengetahui kecenderungan konsumen dalam membeli barang. Metode yang digunakan dalam penelitian ini adalah data mining algoritma apriori pada data penjualan Barbar Warehouse dalam 3 bulan dan diolah menggunakan Rapid Miner. Hasil pengolahan data dengan Rapid Miner menunjukkan bahwa penjualan produk yang paling banyak terjual yaitu pada kategori beding (seprei) dan watches (kacamata).

2.2. Kerangka Pemikiran



Gambar 2.3 Kerangka Pemikiran

Uraian dari kerangka pemikiran penelitian tugas akhir dapat ditunjukkan pada Gambar 2.3.

- a. Pokok permasalahan yang mendasari perlunya menerapkan data mining pada transaksi penjualan pulsa menggunakan algoritma apriori.
- b. Perumusan masalah merupakan inti permasalahan dan jalan keluar menyelesaikan permasalahan.
- c. Pengumpulan data tertulis dan tidak tertulis, maksudnya adalah ada penelitian dilakukan pengumpulan data secara tertulis dan tidak tertulis pada Counter Gemilang cell. Pengumpulan data penelitian ini menggunakan metode observasi, dokumentasi, dan wawancara.
- d. Penjualan pulsa, maksudnya adalah hasil atau data dari hasil transaksi penjualan pulsa.
- e. Laporan penjualan pulsa, maksudnya adalah laporan hasil dari beberapa hasil transaksi penjualan pulsa.
- f. Langkah penelitian maksudnya langkah-langkah atau step yang dilakukan peneliti untuk mendapatkan data yang mau di terapkan ke data mining dengan algoritma apriori.
- g. Hasil penerapan data mining pada transaksi penjualan pulsa menggunakan algoritma apriori maksudnya adalah hasil yang didapatkan setelah menerapkan data mining dengan algoritma apriori tersebut.
- h. Algoritma apriori dan aturan asosiasinya maksudnya adalah bagaimana data tersebut diterapkan ke dalam algoritma apriori, dan menentukan aturan asosiasinya.
- i. Data yang dihasilkan maksudnya adalah data yang dihasilkan dari algoritma apriori tersebut.

2.3. Teori Pendukung

2.3.1. Data Mining

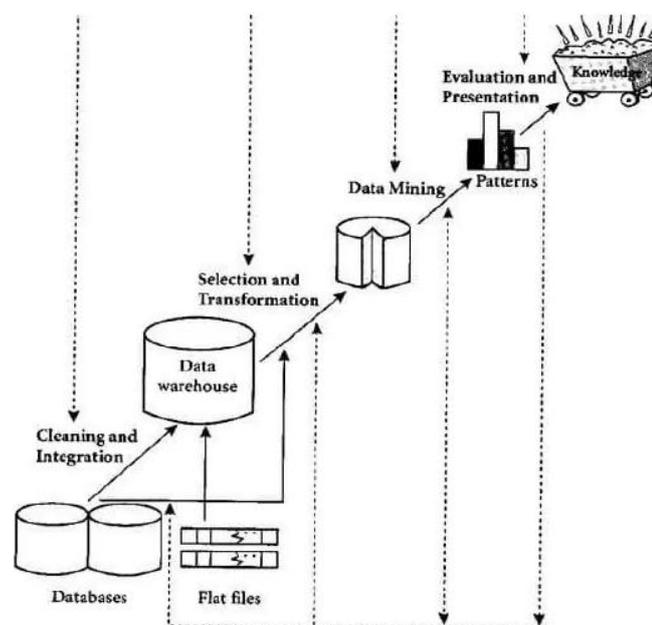
Menurut Indah Werdiningsih., et.al., (2017) Data mining merupakan bidang ilmu yang digunakan untuk menangani masalah pengambilan informasi dari database yang besar dengan menggabungkan Teknik dari statistika, pembelajaran mesin, visualisasi data, pengenalan pola dan database.

Tujuan dari data mining adalah untuk mengekstrak informasi dengan metode cerdas dari kumpulan data kemudian mengubah informasi menjadi struktur yang dapat dipahami untuk penggunaan lebih lanjut.

Sistem terdiri dari tiga tahapan, yaitu input, proses dan output. Pada umumnya, suatu hal diawali dengan input data kemudian dilakukan proses hingga dihasilkan suatu *output*. Fase tersebut juga terdapat dalam data mining. Hasil yang menjadi pembeda pada data mining adalah input berupa set data, prosesnya adalah algoritma atau metode dalam data mining, serta outputnya berbentuk pola, cluster, decision tree dan lain-lain.

Knowledge Discovery Database(KDD)

Menurut Siti Nurhajiza (2019) keseluruhan proses nontrivial untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data.



Gambar 2.4 Proses *Knowledge Discovery Database(KDD)*

(Sumber: Techsi Vol. 6 No.1, April 2019)

Tahapan dalam penggunaan data mining yang merupakan proses *Knowledge Discovery Database (KDD)* dapat diuraikan sebagai berikut:

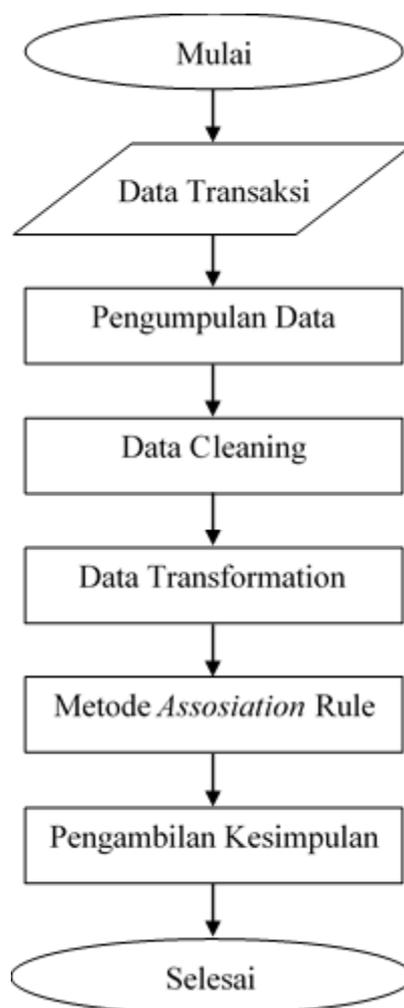
- a. *Data Cleansing*, Proses dimana data-data yang tidak lengkap, mengandung error dan tidak konsisten dibuang dari koleksi data. Ketahui juga data *lifecycle management* untuk mengetahui tentang pengolahan data.
- b. *Data Integration*, Proses integrasi data dimana yang berulang akan dikombinasikan.
- c. *Selection*, Proses seleksi atau pemilihan data yang relevan terhadap analisis untuk diterima dari koleksi data yang ada.
- d. *Data Transformation*, Proses transformasi data yang sudah dipilih ke dalam bentuk mining procedure melalui cara dan agresi data.
- e. *Data Mining*, Proses yang paling penting dimana akan dilakukan berbagai teknik yang diaplikasikan untuk mengekstrak berbagai pola-pola potensial untuk mendapatkan data yang berguna.
- f. *Pattern Evolution*, Sebuah proses dimana pola-pola menarik yang sebelumnya sudah ditemukan dengan identifikasi berdasarkan measure yang telah diberikan.
- g. *Knowledge Presentation*, Merupakan proses tahap terakhir, Dalam hal ini digunakan teknik visualisasi yang bertujuan membantu user dalam mengerti dan menginterpretasikan hasil dari penambangan data.

2.3.2. Algoritma Apriori

Algoritma apriori adalah salah satu algoritma dalam data mining paling terkenal dalam menemukan pola data atau kemunculan/frekuensi data. Algoritma apriori digunakan agar komputer dapat mempelajari aturan asosiasi, mencari pola hubungan antar satu atau lebih item dalam suatu dataset.

Algoritma apriori banyak digunakan pada data transaksi atau biasa disebut market basket, misalnya sebuah swalayan memiliki market basket, dengan adanya algoritma apriori, pemilik swalayan dapat mengetahui pola pembelian seorang konsumen, jika seorang konsumen membeli item A , B, punya kemungkinan 50% dia akan membeli item C, pola ini sangat signifikan dengan adanya data transaksi selama ini.

Tetapi di lain pihak Apriori memiliki kelemahan karena harus melakukan scan database setiap kali iterasi, sehingga waktu yang diperlukan bertambah dengan makin banyak iterasi. Masalah ini yang dipecahkan oleh algoritma-algoritma baru seperti FP-growth.



Gambar 3.2 Flowchart Algoritma Apriori

(Sumber Arlinda Amalia Dewayanti Seminar Nasional Matematika dan Pendidikan Matematika 2018)

2.3.3. Association Rule

Menurut Sari (2017) *Association rule* merupakan salah satu bentuk pola yang dihasilkan oleh Data Mining. *Association rule* dapat digunakan untuk menemukan hubungan atau sebab akibat. Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolak ukur, yaitu : *support* dan *confidence*. *Support* (nilai penunjang) adalah persentasi kombinasi item tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar item dalam aturan asosiasi .

Association rule mempunyai bentuk LHS (Left Hand Shake) -> RHS (Right Hand Shake) dengan interpretasi jika setiap item dalam LHS dibeli maka sepertinya item dalam RHS juga dibeli. Dua pengukuran penting untuk sebuah rule adalah Support dan Confidence. Support (s) didefinisikan sebagai presentasi record terhadap total record keseluruhan dalam database. Support dapat ditulis dalam rumus berikut:

$$\frac{\text{Support banyak}(x)}{\text{Jumlah total transaksi dalam database}} = \dots\dots(1)$$

Sementara nilai support dari 2 item dituliskan dalam rumus berikut:

$$\frac{\text{Support}(X,Y)=P(X \cap Y)}{\sum \text{Banyaknya transaksi X dan Y}} = \dots\dots(2)$$

Jadi jika misalnya Support sebuah item adalah 0,1 % maknanya hanya 0,1persen dari transaksi berisikan item ini. Sementara Confidence (c) didefinisikan sebagai presentasi dari sejumlah transaksi yang berisikan $X \cup Y$

terhadap jumlah record keseluruhan yang mengandung X. Confidence dapat dituliskan dalam rumus berikut :

$$Confidence(X|Y) = \frac{Support(XY)}{Support(X)}$$

Confidence merupakan ukuran kekuatan sebuah rule asosiasi. Jika kita anggap confidence X \rightarrow Y adalah 80%, berarti sebanyak 80 persen transaksi yang mengandung X maka akan mengandung jika Y.

Pada umumnya, menurut Han dan Kamber (2012) association rule mining dapat dilihat sebagai proses dua langkah, yaitu:

- a. Temukan semua *frequent* itemsets, dengan pengertian setiap itemset ini akan muncul sesering sekurangnya minimum *support* (min-sup) yang telah ditentukan.
- b. Hasilkan *association rules* yang kuat dari *frequent* itemsets: dengan pengertian rule-rule ini haruslah memenuhi minimum *Support* dan minimum *Confidence*.

Menurut Eko W. Y (2008) proses aturan asosiasi terdiri dari beberapa tahap sebagai berikut Sistem men-scan database untuk mendapatkan kandidat 1-itemset (himpunan item yang terdiri dari 1 item) dan menghitung nilai supportnya. Kemudian nilai supportnya tersebut dibandingkan dengan minimum support yang telah ditentukan, jika nilainya lebih besar atau sama dengan minimum support maka itemset tersebut termasuk dalam large itemset.

- a. Itemset yang tidak termasuk dalam large itemset tidak diikuti dalam iterasi selanjutnya (diprune).
- b. Pada iterasi kedua sistem akan menggunakan hasil large itemset pada iterasi pertama (L1) untuk membentuk kandidat itemset kedua (L2). Pada iterasi selanjutnya sistem akan menggunakan hasil large itemset pada iterasi selanjutnya akan menggunakan hasil large itemset pada iterasi sebelumnya (Lk-1) untuk membentuk kandidat itemset berikut (Lk). Sistem akan menggabungkan (join) Lk-1 dengan Lk-1 untuk

- mendapatkan L_k , seperti pada iterasi sebelumnya sistem akan menghapus (prune) kombinasi itemset yang tidak termasuk dalam large itemset.
- c. Setelah dilakukan operasi join, maka pasangan itemset baru hasil proses join tersebut dihitung supportnya.
 - d. Proses pembentuk kandidat yang terdiri dari proses join dan prune akan terus dilakukan hingga himpunan kandidat itemsetnya null, atau sudah tidak ada lagi kandidat yang akan dibentuk.
 - e. Setelah itu, dari hasil frequent itemset tersebut dibentuk association rule yang memenuhi nilai support dan confidence yang telah ditentukan.
 - f. Pada pembentukan association rule, nilai yang sama dianggap sebagai satu nilai.
 - g. *Association rule* yang terbentuk harus memenuhi nilai minimum yang telah ditentukan.
 - h. Untuk setiap large itemset L , kita cari himpunan bagian L yang tidak kosong. Untuk setiap himpunan bagian tersebut, dihasilkan rule dengan bentuk $aB(L-a)$ jika supportnya (L) dan supportnya (a) lebih besar dari minimum *support*.

Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, support (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan confidence (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif. Aturan asosiatif biasanya dinyatakan dalam bentuk : {pulsa, kuota} -> {token listrik} (support = 40%, confidence = 50%) Yang artinya : “Seorang konsumen yang membeli pulsa dan kuota punya kemungkinan 50% untuk juga membeli token listrik. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini.” Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk support (minimum support) dan syarat minimum untuk *confidence* (minimum confidence).

