

BAB II

TINJAUAN LITERATUR

2.1 Penelitian Terkait

Dalam penelitian ini untuk membantu dalam proses penelitian dilakukan pengumpulan referensi materi mulai dari penelitian terkait dan materi yang berkaitan dengan judul penelitian ini.

Pertama penelitian dengan judul “Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap”, penelitian ini berfokus pada penggunaan algoritma apriori sebagai metode untuk menentukan persediaan barang pada Toko Sinar Harahap. Masalah dari penelitian ini yaitu toko sering mengalami kekosongan pada persediaan beberapa barang dan kebutuhan yang di cari oleh pelanggan, ini di akibatkan dari tidak adanya kebiasaan pengontrolan persediaan pada toko. Tujuan penelitian ini untuk melihat barang dan kebutuhan apa saja yang dibutuhkan oleh pelanggan toko. Berdasarkan hasil penelitian ini maka di dapat permintaan tertinggi terjadi pada pembelian beras dan juga pembelian telur (Putri, 2022).

Kedua penelitian dengan judul “Penerapan Data Mining Dengan Metode Algoritma Apriori Untuk Menentukan Pola Pembelian Ikan”, penelitian ini berfokus pada penggunaan algoritman apriori untuk membantu dalam menentukan pola pembelian ikan. Masalah dalam penelitian ini yaitu kegiatan transaksi yang berhubungan dengan penjualan dan pembelian masih menggunakan buku catatan seperti untuk pendataan ikan, pendataan penjualan ikan, dan masih memperkirakan ikan yang akan dibeli untuk memenuhi persediaan ikan pada gudang, selain itu masih belum adanya model perhitungan untuk menentukan

belanja persediaan ikan yang paling diminati oleh konsumen. Tujuan penelitian ini untuk membangun aplikasi dengan memanfaatkan algoritma apriori untuk menentukan pola pembelian ikan. Hasil dari penelitian ini telah dibuatnya sebuah sistem untuk menentukan jenis ikan yang paling diminati (Septian, 2019).

Ketiga penelitian dengan judul “Penerapan Algoritma Apriori Untuk Mencari Pola Penjualan Produk Dana Pada PT Bank Rakyat Indonesia (Persero) TBL Kanca Jakarta Pasar Minggu”, penelitian ini berfokus pada penggunaan algoritma apriori untuk menentukan pola penjualan produk dana bank. Masalah dalam penelitian ini yaitu terkait dengan data penjualan produk dana hanya diarsipkan saja. Tujuan penelitian ini untuk membantu menentukan produk dana yang paling diminati nasabah bank. Hasil penelitian dari perhitungan algoritma apriori dapat ditemukan hubungan antar tiap produk dana yang diminati nasabah secara bersamaan. produk dana simpanan yang paling banyak terjual yaitu Britama Rencana dan Britama, dengan menghasilkan kesimpulan bahwa produk dana simpanan Britama Rencana dan Britama dengan nilai *Support* 41,67% dan nilai *Confidence* 71,42% atau produk dana simpanan Britama dan Britama Rencana dengan nilai *Support* 41,67% dan nilai *Confidence* 62,50% (Sutrisno, 2020).

Berdasarkan ketiga penelitian tersebut, acuan yang dapat diterapkan dalam penelitian ini mencakup identifikasi masalah spesifik yang dihadapi oleh Warung Dini terkait pengelolaan persediaan dan pencatatan transaksi penjualan. Kemudian juga dapat menggunakan algoritma Apriori untuk menganalisis pola pembelian produk di Warung Dini dan menentukan produk yang paling diminati oleh pelanggan. Dengan demikian, penelitian ini dapat memberikan solusi praktis

untuk meningkatkan efisiensi pengelolaan persediaan dan penjualan di Warung Dini, serta memastikan ketersediaan produk yang sesuai dengan kebutuhan pelanggan.

Kemudian untuk penelitian terdahulu memiliki perbedaan dengan penelitian ini yaitu ketiga penelitian terdahulu memiliki fokus dan objek yang berbeda dibandingkan dengan penelitian ini. Penelitian Putri (2022) berfokus pada pengendalian persediaan barang di Toko Sinar Harahap, dengan tujuan utama untuk mengetahui barang dan kebutuhan yang paling sering dicari oleh pelanggan guna mengatasi masalah kekosongan persediaan. Penelitian Septian (2019) menyoroti masalah pencatatan manual transaksi dan ketidakmampuan untuk memperkirakan persediaan ikan yang diminati konsumen, dengan tujuan membangun aplikasi menggunakan algoritma Apriori untuk menentukan pola pembelian ikan. Sementara itu, penelitian Sutrisno (2020) memusatkan perhatian pada analisis data penjualan produk dana bank yang selama ini hanya diarsipkan, dengan tujuan untuk menentukan produk dana yang paling diminati oleh nasabah.

Di sisi lain, penelitian ini berfokus pada penerapan algoritma Apriori untuk mencari pola penjualan produk di Warung Dini. Tujuan utama penelitian ini adalah untuk menganalisis pola pembelian produk dan menentukan produk yang paling diminati oleh pelanggan di Warung Dini, sehingga dapat membantu dalam pengelolaan persediaan dan penjualan. Perbedaan utama terletak pada konteks dan objek penelitian, di mana penelitian ini spesifik pada lingkungan warung kecil dengan berbagai macam produk yang dijual, sementara penelitian terdahulu berfokus pada toko bahan pokok, penjualan ikan, dan produk dana bank.

2.2 Pola Penjualan

Menurut Kamus Besar Bahasa Indonesia (2002) pola adalah suatu *system* kerja, sedangkan menurut kamus antropologi pola adalah rangkaian unsur yang sudah mantap mengenai suatu gejala dan dapat dipakai sebagai contoh dalam menggambarkan atau mendeskripsikan gejala itu sendiri. Berdasarkan uraian tersebut, dapat disimpulkan bahwa pola adalah cara kerja yang terdiri dari unsur-unsur terhadap suatu perilaku dan dapat dipakai untuk menggambarkan atau mendeskripsikan gejala perilaku itu sendiri.

Sedangkan Penjualan adalah suatu usaha yang terpadu yang bertujuan untuk mengembangkan rencana-rencana yang diarahkan untuk pemuasan dan kebutuhan serta keinginan pembeli yang menghasilkan laba penjualan (M Ahmadar, 2021).

Jadi, pola penjualan adalah tren atau kecenderungan dalam cara produk atau layanan terjual dalam jangka waktu tertentu. Ini mencakup berbagai aspek seperti volume penjualan, frekuensi, waktu puncak penjualan, preferensi pelanggan, dan lain sebagainya. Memahami pola penjualan dapat membantu bisnis dalam perencanaan stok, strategi pemasaran, dan pengambilan keputusan lainnya.

2.3 Data Mining

Istilah Data Mining sebenarnya mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang dianggap penting dalam berbagai bidang, mulai dari bidang akademik, bisnis, hingga bidang medis. Munculnya data mining didasarkan pada jumlah data yang tersimpan dalam basis data semakin besar. Perkembangan yang cepat dalam teknologi pengumpulan dan penyimpanan data telah memudahkan organisasi untuk mengumpulkan sejumlah data berukuran besar sehingga menghasilkan gunung data (Ketut, 2022).

Data mining adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika. Data mining merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Yuli, 2019).

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.

Proses data mining secara garis besar dapat dijelaskan sebagai berikut :

1. Data Selection Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discovery in Database* (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas terpisah dari basis data operasional.
2. Pre-processing / *Cleaning* Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus *Knowledge Discovery in Database* (KDD). Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan

cetak. Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk *Knowledge Discovery in Database* (KDD), seperti data atau informasi eksternal lainnya yang diperlukan.

3. *Transformation* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam *Knowledge Discovery in Database* (KDD) merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
4. Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses Knowledge Discovery in Database (KDD) secara keseluruhan.
5. *Interpretation / Evaluation* Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses Knowledge Discovery in Database (KDD) yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.4 Aturan Asosiasi (*Association Rule*)

Association rules merupakan sebuah aturan tertentu atau rules yang menyatakan sebuah hubungan korelasi antara tingkat kemunculan beberapa atribut dalam sebuah database. Association Rule adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Karena awalnya berasal dari studi tentang database transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersamaan dengan produk apa, maka aturan asosiasi juga sering dinamakan *Market basket analysis* (M Ikbal, 2015).

Analisis asosiasi atau *association rule mining* adalah teknik data mining untuk menemukan aturan asosiasi antara suatu kombinasi item. Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari salah satu teknik data mining lainnya. Bila kita mengambil contoh aturan asosiatif dalam suatu transaksi pembelian barang disuatu minimarket adalah kita dapat mengetahui berapa besar kemungkinan seorang konsumen membeli suatu item bersamaan dengan item lainnya (seperti membeli roti bersama dengan selai) (Delima, 2023).

1. Analisis Pola Frekuensi Tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* (penunjang) yaitu persentase item atau kombinasi item yang ada pada *database*. Nilai *support* sebuah item diperoleh dengan rumus berikut :

$$Support(A) = \frac{Jumlah\ Transaksi\ Mengandung\ A}{Total\ Transaksi} \times 100\%$$

Rumus support tersebut menjelaskan bahwa nilai *support* didapat dengan cara membagi jumlah transaksi yang mengandung item A (satu item) dengan

jumlah total seluruh transaksi. Sedangkan nilai *support* dari 2 item diperoleh dari rumus berikut:

$$Support (A, B) = P (A \cap B)$$

$$Support (A, B) = \frac{Jumlah\ Transaksi\ Mengandung\ A\ Dan\ B}{Total\ Transaksi} \times 100\%$$

Rumus *support* diatas menjelaskan bahwa nilai *support* 2-itemsets didapat dengan cara membagi jumlah transaksi yang mengandung *item* A dan *item* B (item pertama bersamaan dengan item yang lain) dengan jumlah total seluruh transaksi.

2. Pembentukan Aturan Asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung nilai *confidence* aturan assosiatif A B. Nilai *confidence* dari aturan A B diperoleh dari

rumus sebagai berikut :

$$Confidance = P (B | A) \frac{Jumlah\ Transaksi\ Mengandung\ A\ Dan\ B}{Total\ Transaksi\ Mengandung\ A} \times 100\%$$

Rumus diatas menjelaskan bahwa nilai *confidence* diperoleh dengan cara membagi jumlah transaksi yang mengandung *item* A dan *item* B (*item* pertama bersamaan dengan item yang lain) dengan jumlah transaksi yang mengandung *item* A (*item* pertama atau *item* yang ada disebelah kiri). Untuk menentukan aturan asosiasi yang akan dipilih maka harus diurutkan berdasarkan *Support* × *Confidence*. Aturan diambil sebanyak n aturan yang memiliki hasil terbesar.

2.5 Algoritma Apriori

Algoritma a-priori adalah suatu algoritma dasar yang diusulkan oleh Agrawal dan Srikant pada tahun 1994 untuk menentukan Frequent itemsets untuk aturan asosiasi Boolean. Algoritma A-priori termasuk jenis Aturan Asosiasi pada data mining aturan yang menyatakan asosiasi antara beberapa atribut sering disebut affinity analysis atau market basket analysis (Haryo, 2019)

Perancangan pada algoritma apriori merupakan Langkah-langkah atau *step by step* dari proses algoritma apriori dimana pada langkah-langkah terdiri dari input, proses dan output. Berikut adalah langkah – langkahnya (M Ghofur, 2020):

1. Input
 - a. Data transaksi penjualan yang diambil.
 - b. Tentukan minimal support dan minimal confidence. Dimana minimal support dan minimal confidence akan digunakan sebagai nilai threshold.
2. Proses
 - a. Menghitung iterasi-1 : hitung semua item untuk mendapatkan nilai support (transaksi yang memuat seluruh item) dengan cara men-scan database. Setelah 1-itemset didapatkan, dari 1-itemset apakah diatas atau sama dengan minimal support. Apabila telah memenuhi syarat minimal support maka item tersebut dinyatakan lolos dan akan menjadi pola frekuensi tinggi.
 - b. Menghitung iterasi-2 : untuk mendapatkan 2-itemset dilakukan dengan cara mengkombinasikan dari 1- itemset yang telah dinyatakan lolos pada proses sebelumnya. Kemudian dilakukan

proses untuk mendapatkan nilai support dengan cara men-scan database untuk mendapatkan nilai support. Apabila minimal support dari 2-itemset memenuhi syarat maka dinyatakan lolos.

- c. Menghitung iterasi-3 : untuk mendapatkan 3-itemset dilakukan kombinasi dari 2-itemset yang telah dinyatakan lolos sebagai pola frekuensi tinggi. Setelah dilakukan kombinasi dihitung kembali nilai support. Kemudian dari minimal support yang didapat jika lebih besar dari nilai threshold maka dinyatakan lolos.
- d. Setelah semua frekuensi tinggi ditemukan, maka mencari nilai confidence dari hasil iterasi-2 dan iterasi3 yang memenuhi syarat atau telah dinyatakan lolos.
- e. Hitung nilai confidence dari 3-itemset dan 2-itemset yang telah dinyatakan lolos sebagai pola frekuensi tinggi. Untuk rumus mencari nilai *confidence*.
- f. Apabila nilai confidence lebih besar atau sama dengan nilai minimal confidence maka item tersebut memenuhi syarat atau dinyatakan lolos.

3. Output

Pembentukan aturan assosiatif dari kombinasi 2-itemset dan 3-itemset yang telah dinyatakan lolos.

Berikut merupakan contoh simulasi dari algoritma apriori. Pada simulasi ini data yang digunakan hanyalah tujuh sampel data dan bukan seluruh data. Berikut ini adalah data yang akan digunakan pada simulasi :

Tabel 2.1 Data Transaksi

Transaksi	Item
1	Panci, Baskom , Gayung, Set Rantang, Wajan
2	Sutil, Baskom, Set Rantang , Wajan
3	Tumbler, Set Rantang, Gayung
4	Wajan, Sapu, Kain Pel,
5	Toples, Wajan, Pisau
6	Gayung, Panci, Sapu
7	Kompor, Garpu , Sapu, Kain Pel, Baskom,Set Rantag,Wajan

Selanjutnya apabila dibuat ke dalam bentuk tabular, data transaksi diatas akan terlihat seperti tabel.

Tabel 2.2 Bentuk Tabular Data Transakasi

Item	Transaksi							Σ
	1	2	3	4	5	6	7	
Panci (A)	1	0	0	0	0	1	0	2
Baskom (B)	1	1	0	0	0	0	1	3
Gayung (C)	1	0	1	0	0	1	0	3
Sutil (D)	1	1	1	0	0	0	1	4
Wajan (E)	1	1	0	1	1	0	1	5
Set Rantang (F)	0	1	0	0	0	0	0	1
Tumbler (G)	0	0	1	0	0	0	0	1
Sapu (H)	0	0	0	1	0	1	1	3
Kain Pel (I)	0	0	0	1	0	0	1	2
pisau (J)	0	0	0	0	1	0	0	1
Toples (K)	0	0	0	0	1	0	0	1
Kompor (L)	0	0	0	0	0	0	1	1
Garpu (M)	0	0	0	0	0	0	1	1

Berikut ini adalah pembentukan 1 *itemset* berdasarkan bentuk tabular data transaksi diatas. Dimana pembentukkan 1 *itemset* bertujuan untuk mengetahui nilai *support* dari masing-masing item dalam transaksi yang ada.

Tabel 2.3 Hasil Pembentukan 1 *Itemset*

Nama	Jumlah	Support
Panci (A)	2	28,57%
Baskom (B)	3	42,86%
Gayung (C)	3	42,86%
Sutil (D)	4	57,14%
Wajan (E)	5	71,43%
Set Rantang (F)	1	14,29%
Tumbler (G)	1	14,29%
Sapu (H)	3	42,86%
Kain Pel (I)	2	28,57%
pisau (J)	1	42,86%
Toples (K)	1	14,29%
Kompor (L)	1	14,29%
Garpu (M)	1	14,29%

Berdasarkan tabel hasil pembentukan 1 *itemset* maka penulis menentukan nilai *minimum support* yang digunakan adalah 40%, maka item-item seperti Gayung (C), Sutil (D), Wajan (E), dan Pisau (J) akan termasuk dalam frequent itemset karena memiliki support di atas 40%. Pemilihan nilai minimum support ini dapat disesuaikan dengan eksplorasi lebih lanjut terhadap data dan tujuan analisis yang ingin dicapai.

Berdasarkan hasil pembentukan 1 *itemset* diatas terdapat 4 item yang lebih besar dari nilai *minimum support*. Maka himpunan dari kombinasi 2 *itemset* yang mungkin terbentuk adalah sebagai berikut:

Tabel 2.4 Tabel Calon 2 *Itemset*

Kombinasi Itemset	Frekuensi (f)
{C,D}	3
{C,E}	4
{C,J}	2
{D,E}	5
{D,J}	2
{E,J}	4

Berdasarkan bentuk tabular data calon 2 *itemset* diatas, dilakukan pencarian nilai *support* dari 2 *itemset*. Pencarian nilai *support* 2 *itemset* bertujuan untuk mengetahui frekuensi kemunculan 2 *itemset* dibeli secara bersamaan.

Tabel 2.5 Hasil nilai *Support* 2 *itemset*

Kombinasi Itemset	Frekuensi (f)	Support (%)
{C,D}	1	14.29%
{C,E}	2	28.57%
{C,J}	0	0.00%
{D,E}	3	42.86%
{D,J}	0	0.00%
{E,J}	1	14.29%

Telah dijelaskan oleh penulis bahwa nilai minimum support yang digunakan adalah 40%, maka kombinasi yang memenuhi nilai minimum support adalah {D,E} karena nilai support-nya 42.86% atau lebih dari 40%.

Berdasarkan kombinasi 2 *itemset* yang terbentuk maka didapat aturan hubungan yang ada yaitu :

1. Jika membeli D maka membeli E
2. Jika membeli E maka membeli D

Selanjutnya menghitung nilai *support* dan *confidence*. Berikut ini adalah perhitungan nilai *support* dan *confidence* dari aturan yang didapat.

Tabel 2.6 Perhitungan nilai *Support* dan *Confidence*

Aturan	Support		Confidence	
Jika membeli D maka membeli E	$(3/7) \times 100\% =$	42.86%	$(4/3) \times 100\% =$	75%
Jika membeli E maka membeli D	$(3/7) \times 100\% =$	42.86%	$(5/3) \times 100\% =$	60%

Setelah didapat nilai *support* dan *confidence* untuk masing-masing kandidat, dilakukan perkalian antar *support* dan *confidence*, dimana kandidat yang diambil adalah kandidat yang memiliki nilai *support* $\geq 40\%$ dan *confidence* $\geq 60\%$, sehingga didapat tabel sebagai berikut:

Tabel 2.7 Kandidat yang Mencapai Nilai *Support* dan *Confidence*

Aturan	Support (%)	Confidence (%)	Support $\times$ Confidence
Jika membeli D maka membeli E	42.86%	75%	32.14
Jika membeli E maka membeli D	42.86%	60%	25.71

Setelah didapat hasil perkalian antar *support* dan *confidence*, maka akan dipilih yang hasil perkaliannya paling besar karena perkalian tersebut merupakan *rule* yang digunakan saat penjualan. Namun, karena pada hasil perkalian diatas bernilai samamaka semuanya bisa dijadikan *rule*, yaitu :

1. “Jika membeli Sutil (D) maka membeli Wajan (E) dengan *support* 42.86% dan *confidence* 75%”.
2. “Jika membeli Wajan (E) maka membeli Sutil (D) dengan *support* 42.86% dan *confidence* 60%”.

Dari dua aturan diatas dapat ditarik kesimpulan apabila melakukan persediaan Sutil maka lakukan persediaan juga terhadap Wajan.