
PENERAPAN DATA MINING UNTUK MENENTUKAN POLA PEMBELIAN PRODUK DENGAN ALOGRITMA APRIORI

Tintin Chandra

Fakultas Ilmu Sains dan Teknologi

Universitas IBBI

Jalan Sei Deli No. 18 Medan 20114

email: tinuhnsbm@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengidentifikasi pola pembelian produk yang sering terjadi secara bersamaan oleh konsumen pada bisnis produk. Teknik Data Mining dipilih pada penelitian Adalah data transaksi penjualan untuk periode tertentu Metode yang dipilih dalam penelitian adalah Algoritma Apriori untuk menghasilkan produk yang sering dipesan atau dibeli oleh konsumen. Parameter yang dipilih adalah Minimum Support (Min. S) sebesar 3% dan Minimum Confidence (Min. C) sebesar 70%. Hasil analisis memberikan informasi adanya beberapa pola aturan asosiasi yang kuat, bahwa konsumen yang membeli "produk A " cenderung membeli "produk B" dengan tingkat kepercayaan 85%. Hasil analisis menunjukkan bahwa wawasan strategis untuk penempatan produk, dan pengambilan keputusan bisnis lainnya dapat meningkatkan efisiensi dan pengelolaan stok produk.

Kata Kunci: Data Mining, Algoritma Apriori, Aturan Asosiasi, Market Basket Analysis.

1. Pendahuluan

Para pelaku bisnis khususnya perintis tentu ingin agar produk yang ditawarkan laris manis di masyarakat. Pelaku bisnis yang menjalankan usahanya tentu harus mengetahui data penjualan setiap hari, setiap minggu, sampai setiap bulan. Untuk itu para pelaku bisnis harus mencermati pola pembelian yang dilakukan oleh konsumen. Pelaku bisnis wajib mengetahui data penjualan dimana data penjualan berguna untuk menentukan strategi penjualan yang tepat, menganalisis tingkat penjualan setiap bulan. Hasil analisis data membantu para pelaku bisnis untuk mengetahui seberapa banyak barang yang restock agar konsumen tidak pergi ke pesaing untuk membeli produk yang dicari. Apabila data penjualan lengkap maka pelaku bisnis dapat memantau perkembangan bisnis secara tepat waktu, bahkan pelaku bisnis dapat memahami produk mana yang paling diminati konsumen. Peneliti melihat banyak kekurangan yang terjadi pada pelaku bisnis, dimana permasalahan paling umum Adalah menjual produk yang kurang sesuai dengan perilaku kebiasaan konsumen dalam membeli produk. Pembelian yang dilakukan konsumen akan mempengaruhi tingkat penjualan barang dimana produk yang dijual kurang diminati oleh konsumen. Data transaksi penjualan dari aneka produk akan terus bertambah setiap harinya yang menyebabkan penyimpanan data semakin besar.

Volume data transaksi penjualan yang ada pada pelaku bisnis modern terus meningkat. Data yang tersimpan dalam database memiliki informasi yang berguna mengenai perilaku dan preferensi customer. Untuk mengolah data mentah menjadi wawasan yang dapat diproses, diperlukan teknik analitik salah satunya dengan menggunakan Data Mining. Penggunaan Data Mining memberikan solusi melalui metode Market Basket Analysis yang dapat diterapkan dengan Algoritma Apriori, dimana algoritma apriori dirancang secara efisien mengekstrak pola produk yang sering dibeli bersama. Dalam konteks sparepart, pola ini dapat menginisialisasi komponen yang dalam pemasangan bisa saling melengkapi (misalnya seal dan packing) atau yang secara fungsi diperlukan dalam satu kali perbaikan (misalnya oli mesin dan oli gardan). Kebanyakan data transaksi penjualan hanya dijadikan arsip saja tanpa dimanfaatkan dengan baik. Padahal kumpulan data tersebut memiliki informasi yang sangat bermanfaat. Berdasarkan permasalahan di atas, maka diperlukan adanya aplikasi untuk mengelompokkan data barang berdasarkan kecenderungannya yang muncul bersamaan dalam suatu transaksi menggunakan algoritma apriori. Penerapan algoritma apriori dalam penelitian ini diharapkan akan menemukan pola berupa produk yang sering dibeli bersamaan. Pola tersebut bisa digunakan untuk menempatkan produk yang sering dibeli bersamaan dalam sebuah area yang saling berdekatan, merancang kupon diskon pada produk tertentu untuk menarik daya beli konsumen. Rumusan Masalah Adapun rumusan masalahnya adalah bagaimana membuat aplikasi data mining analisa pola pembelian produk dengan metode algoritma apriori untuk memberikan gambaran keterkaitan antar barang dengan menganalisis data transaksi penjualan berdasarkan pola berupa produk barang yang sering dibeli secara bersamaan. Tujuan Penulisan Tujuan penulisannya adalah membangun aplikasi penerapan data mining analisa pola pembelian produk dengan metode algoritma apriori untuk menemukan pola pembelian barang yang sering dibeli secara bersamaan guna pengembangan strategi pemasaran dalam penjualan barang.

Penelitian yang diteliti menggunakan data mining untuk menentukan pola pembelian produk dengan algoritma apriori nantinya dapat membantu pelaku usaha yang ingin mengetahui pola pembelian oleh customer dan untuk menjaga stok agar overload atau tidak sampai kosong. Selain itu, penelitian dengan algoritma apriori

diharapkan dapat memberikan dasar dukungan kepada pelaku bisnis dalam mengambil keputusan yang lebih tepat sasaran.

2. Landasan Teori

Data mining merupakan salah satu proses eksplorasi dan analisis data yang memiliki banyak algoritma dengan kegunaan sesuai permasalahan yang ingin diselesaikan. Data mining merupakan gabungan dari berbagai disiplin bidang ilmu antara lain basis data, information retrieval, statistika, machine learning dan sebagainya. Data mining dapat diterapkan pada berbagai bidang seperti bisnis, kesehatan, asuransi, pemasaran, dan maupun yang lainnya yang disesuaikan dengan kebutuhan. Data mining merupakan cara untuk menemukan informasi yang tersembunyi dalam sebuah kondisi dan merupakan bagian dari proses Knowledge Discovery in Database (KDD). Data Mining sering juga disebut knowledge discovery in database (KDD) karena kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Output dari data mining dapat digunakan untuk memperbaiki pengambilan keputusan di masa depan. Data mining merupakan gabungan teori dan heuristik, fokus pada proses penemuan knowledge/pola termasuk data cleaning, learning, dan visualisasi dari hasilnya. Clustering juga dikenal sebagai unsupervised learning yang membagi data menjadi kelompok atau clusters berdasarkan pengelompokan dari data yang dikumpulkan. Karakteristik tiap cluster tidak ditentukan sebelumnya, tetapi tercermin dari kemiripan data yang terkelompok di dalamnya. Teknik clustering yang digunakan data mining meliputi: skalabilitas, kemampuan untuk menangani tipe atribut yang berbeda, menangani data yang mengandung noise, mampu menangani dimensionalitas yang tinggi, dan dapat diterjemahkan dengan mudah.

Karakteristik data mining adalah Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dari pola data tertentu, data mining biasa menggunakan data yang sangat besar, data yang besar digunakan untuk membuat hasil lebih akurat, data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi. Tahapan proses data mining meliputi Pembersihan data (Data cleaning) merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Integrasi data (Data integration) merupakan penggabungan data dari berbagai database ke dalam satu database baru. Seleksi data (Data selection) dimana data yang ada pada database sering kali tidak digunakan semuanya, maka hanya data yang sesuai untuk dianalisis atau yang cocok dengan data uji yang akan diambil dari database. Transformasi data (Data transformation) yaitu Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Proses mining merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Evaluasi pola (Pattern evaluation) Untuk mengidentifikasi pola knowledge based yang ditemukan. Tahapan dari hasil evaluasi data mining berupa pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Presentasi pengetahuan (Knowledge presentation) yaitu visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Algoritma Apriori adalah salah satu bagian dari sebuah metode Association Rule pada data mining yang berfungsi untuk mencari frequent item (produk) pada transaksional database. Association rule mining merupakan teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Algoritma ini menggunakan pendekatan “bottom-up” untuk menentukan produk paling sering muncul. Pendekatan ini menggunakan Breadth-first search dan struktur data Hash tree untuk menghitung kandidat produk secara efisien. Pendekatan bottom up dimulai dari setiap item dalam daftar produk. Algoritma apriori dibagi menjadi beberapa tahap yang disebut narasi. Tahapan dari algoritma apriori Adalah pembentukan kandidat produk dari kombinasi (k-1) produk yang didapat dari iterasi sebelumnya. Cara dari algoritma apriori adalah pemangkasan kandidat k-produk yang subsetnya berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1. Penghitungan support dari tiap kandidat k-produk. Support dari tiap kandidat k-produk didapat dengan men-scan database untuk menghitung jumlah transaksi yang memuat semua item didalam kandidat k-produk tersebut. Tetapkan pola frekuensi tinggi yang memuat k item atau k-produk ditetapkan dari kandidat k-produk yang supportnya lebih besar dari minimum support. Apabila tidak ditemukan pola frekuensi tinggi baru maka seluruh proses dihentikan.

Apriori adalah algoritma yang digunakan untuk menambang Frequent Produk (himpunan item yang sering muncul) dari database transaksi. Algoritma ini bekerja dalam dua fase utama:

1. Generate Frequent Produk: Menggunakan sifat apriori property (menghilangkan produk kandidat yang k-nya lebih besar jika salah satu subset (k-1) tidak lolos ambang batas Min. S.
2. Generate Association Rules: Membuat aturan ($A \rightarrow B$) dari Frequent Produk. Kekuatan aturan diukur menggunakan tiga metrik utama:

1. Support (S): Mengukur seberapa sering produk muncul dalam data.

$$s(A \rightarrow B) = \frac{\text{Jumlah transaksi mengandung } (A \cup B)}{\text{Total Transaksi}}$$

2. Confidence (C): Mengukur seberapa sering aturan terbukti benar.

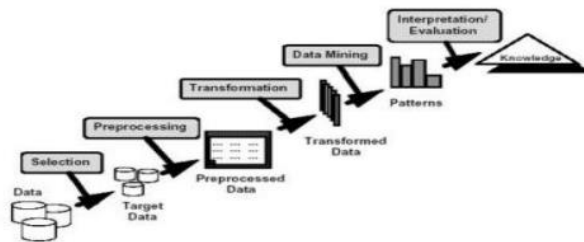
$$C(A \rightarrow B) = \frac{S(A \cup B)}{S(A)}$$

3. Lift (L): Mengukur seberapa besar peningkatan probabilitas pembelian B jika A sudah dibeli, relatif terhadap probabilitas pembelian B saja. Nilai $L > 1$ menunjukkan asosiasi positif.

$$L(A \rightarrow B) = \frac{C(A \rightarrow B)}{S(B)}$$

3. Metode Penelitian

Metode penelitian yang digunakan dalam menganalisis pola pembelian barang dengan menggunakan algoritma apriori yaitu dengan menggunakan tahapan data mining atau Knowledge Discovery in Database (KDD) seperti pada gambar 1.



Gambar 1. Tahapan dalam KDD

Pemilahan data membuat himpunan data target, pengelompokan himpunan data, atau menfokuskan pada subset variabel atau contoh data. Pemilahan data dari kelompok data operasional dilakukan sebelum tahap penggalian informasi. Hasil data pemilahan yang digunakan pada proses data mining. Tahap transaksi hasil pemilahan yang diolah untuk proses data mining disimpan dalam data Excel, terpisah dari basis data utama. Pembersihan data dilakukan seperti penghapusan noise meliputi membuang data ganda, data yang tidak konsisten, data yang tidak dipakai, dan memperbaiki kesalahan pada data. Tahapan hasil transaksi pemilihana yang telah disimpan ke dalam Excel selanjutnya diproses pembersihan melewati tabel yang tidak di butuhkan.

Transformation dalam menemukan fitur yang bermanfaat dalam mempresentasikan data tergantung kepada tujuan yang ingin dicapai. Transformasi data dimana proses data yang telah dipilah yang nantinya data untuk diolah oleh data mining. Proses kreatif tergantung kepada pola informasi yang dicari dalam basis data. Proses menemukan pola atau informasi yang menarik dalam data terpilih dengan menggunakan algoritma tertentu. Pemilihan metode atau algoritma bergantung pada data yang diolah dan tujuan dari proses KDD Evaluation pola yang dihasilkan dari data olahan pada data mining. Pola informasi yang ditampilkan dari proses data mining perlu dibuat dalam bentuk yang mudah dimengerti oleh pelaku bisnis. Tahapan dari proses KDD yang meliputi pemeriksaan apakah pola atau informasi yang dihasilkan berlawanan dengan hipotesa yang ada. Presentasi pengetahuan hasil visualisasi dan penyajian pengetahuan tentang metode yang dapat digunakan untuk menambah pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah memetakan dan memperkuat dalam pengambilan keputusan dari hasil analisis yang diolah.

4. Hasil Penelitian

Database dari transaksi belanja oleh konsumen seperti pada tabel 1.

Tabel 1. Transaks

Transaksi	Produk yang dibeli
1	Susu, Teh, Caramel
2	Teh, Caramel, Biskuit
3	Teh, Caramel
4	Susu, Biskuit
5	Susu, Caramel, Biskuit
6	Teh, Caramel
7	Caramel , Kopi, Susu
8	Caramel , Kopi, Susu
9	Susu, Biskuit, Kopi
10	Caramel, Teh, Kopi

Data pada Tabel 1. menunjukkan tipe produk yang dibeli oleh konsumen dalam periode tertentu dapat direpresentasikan seperti tabel 2.

Tabel 2. Representasi Transaksi

Transaksi	Item yang dibeli
1	Susu
1	Teh
1	Caramel
2	Teh
2	Caramel
2	Biskuit
3	Teh
3	Caramel
4	Susu
4	Biskuit
5	Susu
5	Caramel
5	Biskuit
6	Teh
6	Caramel
7	Caramel
7	Kopi
7	Susu
8	Caramel
8	Kopi
8	Susu
9	Susu
9	Biskuit
9	Kopi
10	Caramel
10	Teh
10	Kopi

Data pada tabel 2. transaksi item yang dibeli dibuat dalam tabularasi seperti pada tabel 3.

Tabel 3. Tabularasi

Transaksi	Teh	Caramel	Kopi	Susu	Biskuit
1	1	1	0	1	0
2	1	1	0	0	1
3	1	1	0	0	0
4	0	0	0	1	1
5	0	1	0	1	1
6	1	1	0	0	0
7	0	1	1	1	0
8	0	1	1	1	0
9	0	0	1	1	1
10	1	1	1	0	0

Calon 2 produk pada setiap data transaksi dan frekuensi masing-masing dihitung dengan data tabular pada tabel 4 yang menunjukkan calon 2 produk dari data transaksi pada tabel 3.

Tabel 4. Tabel calon 2 produk

Kombinasi	Jumlah
Teh, Caramel	5
Teh, kopi	1

Teh, susu	1
Teh, Biskuit	1
Caramel, kopi	3
Caramel, susu	4
Caramel, Biskuit	2
Kopi, susu	3
Kopi, Biskuit	1
Susu, Biskuit	3

Data pada tabel 4 dilanjutkan pada tahap selanjutnya dimana seleksi frekuensi \geq batas minimal yang ditentukan. Jika minimal *supportnya* ditetapkan nilai = 2, maka: $F2 = \{\{\text{teh, Caramel}\}, \{\text{Caramel, kopi}\}, \{\text{Caramel, susu}\}, \{\text{Caramel, Biskuit}\}, \{\text{kopi, susu}\}, \{\text{susu, Biskuit}\}\}$ seperti pada tabel 5.

Tabel 5. Tabel calon 2 produk

Kombinasi	Jumlah
Teh, Caramel	5
Caramel, kopi	3
Caramel, susu	4
Caramel, Biskuit	2
Kopi, susu	3
Susu, Biskuit	3

Kombinasi 2 produk dalam F2 pada tabel 5 dapat di gabungkan menjadi calon 3 produk. Produk dari F2 yang dapat digabungkan adalah produk-produk yang memiliki kesamaan dalam k-2 item pertama. Calon 3 produk yang dapat dibentuk dari F2 tampak pada tabel 6.

Tabel 6. Calon 3 produk

Kombinasi	Jumlah
Teh, Caramel, susu	1
Teh, Caramel, Biskuit	1
Caramel, susu, Biskuit	1
Caramel, kopi, susu	2
Kopi, susu, Biskuit	1
Teh, Caramel, kopi	1

Dengan demikian $F3 = \{\text{Caramel, kopi, susu}\}$, karena hanya kombinasi inilah yang memiliki frekuensi kemunculan ≥ 2 . Pembentukan aturan asosiasi setelah semua pola frekuensi tinggi dipetakan, kemudian cari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan *asosiasi* “ jika A maka B “. Nilai *confidence* dari aturan “ jika A maka B “ Berdasarkan tabel 3, dapat dilihat besarnya nilai *support* dan *confidence* dari calon aturan asosiasi seperti tampak pada tabel 7.

Tabel 7. Aturan Asosiasi 3 produk

Aturan	Confidence	
Jika membeli Caramel dan susu, maka akan membeli kopi	2/4	50%
Jika membeli Caramel dan kopi, maka akan membeli susu	2/3	67%
Jika membeli kopi dan susu, maka akan membeli Caramel	2/3	67%

Misalkan ditetapkan nilai *minimum confidence* 60%, maka aturan yang bisa terbentuk adalah dengan dua aturan:

- Jika membeli Caramel dan kopi, maka akan membeli susu
- Jika membeli kopi dan susu, maka akan membeli Caramel

Calon aturan asosiasi dari F2 dapat dilihat pada tabel 8.

Tabel 8. Aturan Asosiasi 2 produk

Aturan	Confidence	
Jika membeli teh, maka akan membeli Caramel	5/5	100.0%
Jika membeli Caramel, maka akan membeli teh	5/8	62.5%
Jika membeli Caramel, maka akan membeli kopi	3/8	37.5%
Jika membeli kopi, maka akan membeli Caramel	3/4	75.0%
Jika membeli Caramel, maka akan membeli susu	4/8	50.0%
Jika membeli susu, maka akan membeli Caramel	4/6	67.0%
Jika membeli Caramel, maka akan membeli Biskuit	2/8	25.0%
Jika membeli Biskuit, maka akan membeli Caramel	2/4	50.0%
Jika membeli kopi, maka akan membeli susu	3/4	75.0%
Jika membeli susu, maka akan membeli kopi	3/6	50.0%
Jika membeli susu, maka akan membeli Biskuit	3/6	50.0%
Jika membeli Biskuit, maka akan membeli susu	3/4	75.0%

Aturan Asosiasi Final terurut berdasarkan *support x confidence* terbesar seperti pada tabel 9.

Tabel 9. Aturan Asosiasi Final**Tabel 9.** Aturan Asosiasi Final

Aturan	Support	Confidence	Support x Confidence
Jika membeli teh, maka akan membeli caramel	50%	100.0%	50.0%
Jika membeli Caramel, maka akan membeli teh	50%	62.5%	31.3%
Jika membeli susu, maka akan membeli Caramel	40%	67.0%	26.8%
Jika membeli kopi, maka akan membeli Caramel	30%	75.0%	22.5%
Jika membeli kopi, maka akan membeli susu	30%	75.0%	22.5%
Jika membeli Biskuit, maka akan membeli susu	30%	75.0%	22.5%
Jika membeli Caramel dan kopi, maka akan membeli susu	20%	67.0%	13.4%
Jika membeli kopi dan susu, maka akan membeli Caramel	20%	67.0%	13.4%

Tabel *final association rule* menjelaskan tentang *support* dan *confidence* dari masing-masing kombinasi 2 *produks* dan 3 *produks*. Hasil perhitungan *support* pada tabel *final association rule* didapatkan dari jumlah transaksi mengandung A dan B dibagi total transaksi. Untuk hasil *confidence* didapatkan dari jumlah transaksi mengandung A dan B dibagi jumlah transaksi mengandung A. Hasil perkalian *support* dan *confidence* itulah yang menjadi hasil akhir dari algoritma apriori.

5. Kesimpulan

Hasil penelitian yang dilakukan oleh peneliti dalam menerapkan algoritma apriori pada data transaksi usaha bisnis untuk menentukan pola pembelian produk menunjukkan produk mana yang perlu di jaga kesediaan stok. Pola pembelian oleh konsumen memberikan data produk dengan pembelian paling sedikit transaksi dengan label Min. label Min ditemukan beberapa aturan asosiasi yang memiliki Lift signifikan (> 1), memberikan hubungan pembelian yang kuat. Pola yang sangat relevan untuk optimasi tata letak usaha bisnis dan strategi cross-selling. Analisis Apriori dapat diterapkan pada segmen pelanggan tertentu (berdasarkan usia, loyalitas, atau waktu belanja) untuk mendapatkan aturan asosiasi yang lebih spesifik dan personal.

6. Daftar Pustaka

- [1] Jayadi, Patombongi, A., 2017, Implementasi Aplikasi Data Mining pada Apotek Kimia Farma Bahteramas Menggunakan Algoritma Apriori, Jurnal SIMTEK, Vol. 2, No. 1.
- [2] Nofriansyah, D., Erwansyah, K., Ramadhan, M., 2016, Penerapan Data Mining dengan Algoritma Naïve Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan TerhadapKartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi), Jurnal SAINTIKOM, vol.15, No. 2.
- [3] Nurjoko, Kurniawan, H., 2016, Aplikasi Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Apriori di IBI Darm-ajaya Bandar Lampung, Jurnal TIM Darmajaya, vol. 02, No. 01.

- [4] Ristianingrum, Sulastri., 2017, Implementasi Data Mining Menggunakan Algoritma Apriori, Prosiding SINTAK, Vol. 2, No. 2.
- [5] Salamah, E.N., Ulinnuha, N., 2017, Analisis Pola Pembelian Obat dan Alat Kesehatan di Klinik Ibu dan Anak Graha Amani dengan Menggunakan Algoritma Apriori, Jurnal INFORM, Vol.2, No.1.
- [6] Yanto, R., Kesuma, H.D., 2017, Pemanfaatan Data Mining untuk Penempatan Buku di Perpustakaan Menggunakan Metode Association Rule, Jurnal Jatisi, Vol. 4, No.1.