



PENERAPAN DATA MINING UNTUK PREDIKSI POLA PEMBELIAN PELANGGAN MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS: TOKO JIHAN)

Ratna Arista¹, Agung Nugroho², Nanang Tedi Kurniadi³

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

¹ratnaarista@mhs.pelitabangsa.ac.id, ²agung@pelitabangsa.ac.id, ³nanang@pelitabangsa.ac.id

Abstrak

Penentuan kombinasi item dan tata letak barang berdasarkan kecenderungan pembelian konsumen menjadi salah satu solusi bagi Toko Jihan dalam mengembangkan strategi pemasaran sehingga dapat meningkatkan penjualan pada Toko. Algoritma yang dapat digunakan untuk mencari kombinasi item barang apa saja yang sering dibeli secara bersamaan dalam suatu waktu yaitu Algoritma Apriori, Algoritma apriori merupakan algoritma *market basket analysis* yang digunakan untuk menghasilkan *association rule*, dengan pola "if then". Pada algoritma apriori ditentukan *frequent itemset-1*, *frequent itemset-2*, dan *frequent itemset-3* hingga memperoleh aturan asosiasi dari data-data yang sudah diseleksi sebelumnya. Untuk mendapatkan *frequent itemset* tersebut maka setiap data yang sudah diseleksi harus memenuhi syarat *minimum support* dan *minimum confidence*. Pada penelitian ini menggunakan perbandingan *minimum support* dan *minimum confidence* yang berbeda-beda berdasarkan data transaksi yang ada dengan menggunakan *minimum support* 20% dan *minimum confidence* 80% menghasilkan empat aturan asosiasi. Salah satu contohnya yaitu jika konsumen membeli Minyak goreng, Kopi maka 87% (kepastian konsumen dalam membeli *item*) akan membeli Telur.

Kata kunci: *Association Rule Mining, Algoritma Apriori, Support, Confidence.*

Abstract

Determining the combination of items and the layout of goods based on consumer purchasing trends is one solution for Toko Jihan in developing marketing strategies so as to increase sales at the store. The algorithm that can be used to find any combination of items that are often purchased together at a time is the Apriori Algorithm, the apriori algorithm is a market basket analysis algorithm used to generate association rules, with an "if then" pattern. In the apriori algorithm, frequent itemset-1, frequent itemset-2, and frequent itemset-3 are determined to obtain association rules from previously selected data. To get the frequent itemset, each data that has been selected must meet the minimum support and minimum confidence requirements. In this study using different minimum support and minimum confidence comparisons based on existing transaction data using a minimum support of 20% and a minimum confidence of 80% resulted in four association rules. One example is if the consumer buys cooking oil, coffee then 87% (certainty of consumers in buying items) will buy eggs.

Keywords: *Association Rule Mining, Apriori Algorithm, Support, Confidence.*

1. Pendahuluan

Persaingan dalam dunia bisnis yang semakin ketat memaksa para pelaku bisnis untuk berpikir lebih jauh untuk menyusun strategi guna menghadapi persaingan tersebut. Salah satu strategi yang dapat digunakan adalah pemanfaatan teknologi informasi. Dipercayai bahwa teknologi informasi dapat membantu bisnis dan mengelola bisnis. Bisnis dapat memanfaatkan data yang digunakan untuk mendukung pengambilan keputusan yang jika ditangani dengan benar, dapat menghasilkan wawasan yang berharga.

Toko Jihan adalah usaha *retail* yang menjual hampir semua kebutuhan pokok atau sembako, *snack* kemasan, kosmetik, dan lainnya. Setiap harinya Toko Jihan terjadi beberapa transaksi penjualan yang apabila dibiarkan begitu saja hanya dapat menjadi data arsip. Berdasarkan data pembelian konsumen tersebut dapat diolah menjadi sebuah data yang dapat mendukung pengambilan keputusan. Banyaknya data transaksi pembelian yang ada pada Toko Jihan dapat dijadikan

sebagai analisa perilaku pelanggan dalam melakukan pembelian barang. Dalam melakukan pengolahan data tersebut diperlukan sebuah algoritma untuk mengelola data transaksi pembelian konsumen yang ada sehingga dapat menghasilkan pola hubungan antar barang yang dibeli pelanggan.

Dalam penelitian ini menggunakan algoritma apriori yaitu dengan memberikan hubungan antar item dengan data penjualan. Dalam hal ini adalah produk-produk yang dibeli sehingga dapat menghasilkan pola hubungan antar barang yang dibeli pelanggan. Dengan demikian penerapan algoritma apriori, membantu Toko Jihan untuk menemukan pola hubungan *itemset* yang dapat digunakan untuk mengembangkan strategi bisnis.

2. Kajian Pustaka

2.1. Penjualan

Penjualan adalah salah satu syarat berlangsungnya sebuah usaha, dengan penjualan melakukan proses penjualan usaha tersebut akan mendapatkan sebuah keuntungan. Semakin tinggi proses jual beli yang terjadi maka keuntungan yang didapatkan akan berbanding lurus[1].

2.2. Data Mining

Proses pengolahan data menjadi satu hal sangat penting di era persaingan bisnis yang mengharuskan mendapat informasi cepat

Data Mining adalah ilmu yang menggambarkan pengamatan dalam bentuk pengetahuan tentang pengumpulan data untuk menemukan pola dan aturan, pola tersebut didapatkan dari relasi database, data transaksi[2].

2.3. Algoritma Apriori

Algoritma apriori adalah algoritma dari teknik *association rules mining* dan termasuk dari bagian metode data mining. Aturan asosiatif algoritma apriori terbentuk jika-maka[3].

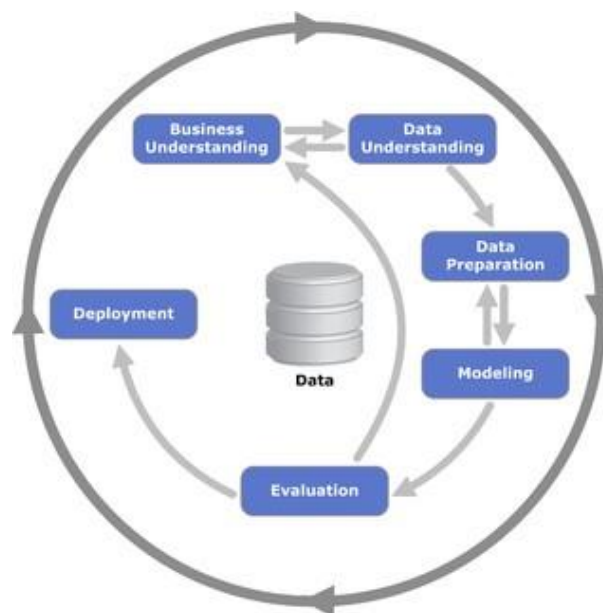
2.4. Association Rule

Association rule bertujuan untuk mencari hubungan atau kekuatan antar item dalam sebuah kombinasi itemset[4] Asosiasi adalah aturan yang mendukung pengolahan data, mengidentifikasi korelasi (generalisasi asosiasi), dan mempertimbangkan urutan pengambilan materi sebagai suatu basis pembangkit aturan.

2.5. CRISP-DM

CRISP-DM merupakan suatu metodologi yang menerapkan pendekatan terstruktur untuk perencanaan proyek data mining yang sudah teruji dengan baik[5]. Model proses CRISP-DM memberikan gambaran tentang siklus hidup proyek data mining.

CRISP DM bukan merupakan satu-satunya standard dalam data mining namun merupakan yang terpopuler saat ini. Berdasarkan hasil pooling *datascience-pm*, CRISP DM Digunakan 2 sampai dengan 3 kali lebih banyak dari 4 teratas standard yang paling banyak digunakan.



Gambar 1 Fase CRISP-DM

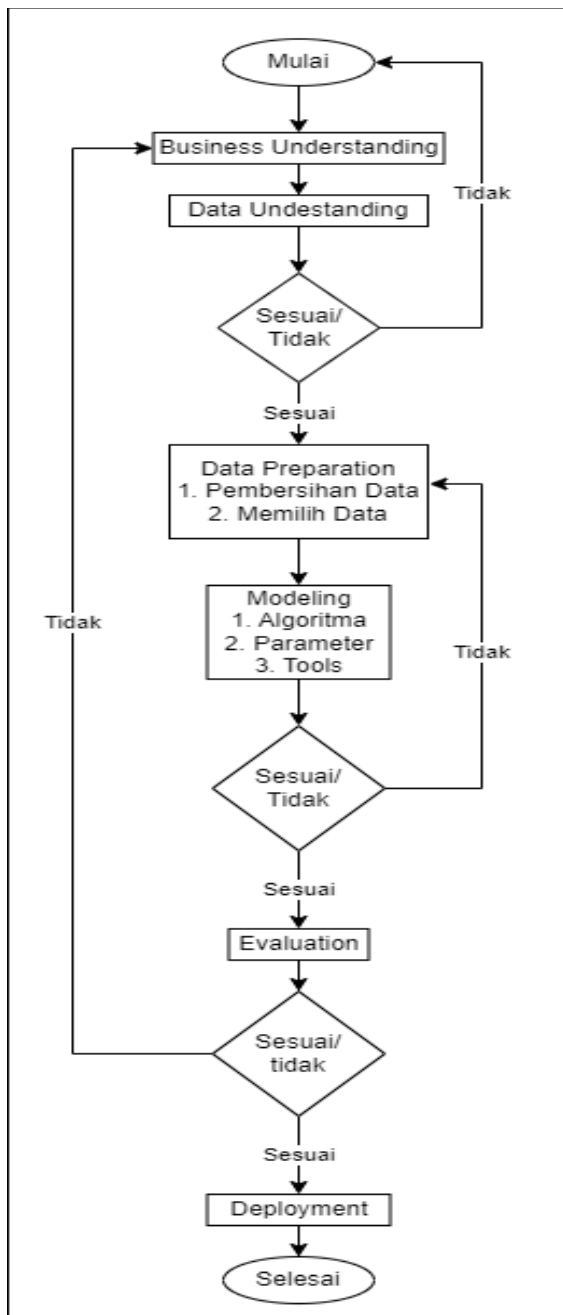
Proses data mining berdasarkan CRISP-DM terdiri dari 6 fase, yaitu[6] :

- Fase pemahaman bisnis
Tahap ini adalah memahami tujuan dan kebutuhan *business*, kemudian mengartikan pengetahuan ke dalam data mining.
- Fase pemahaman data
Tahap pengumpulan data yang kemudian akan di proses untuk mendapatkan pemahaman yang mendalam tentang data.
- Fase persiapan data
Fase ini meliputi semua langkah dalam membangun *dataset* akhir dari data mentah.
- Fase pemodelan
Tahap ini dilakukan pemilihan dan penerapan berbagai teknik dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang baik.
- Fase Evaluasi
Tahap ini dilakukan evaluasi terhadap kinerja dan kualitas model.
- Fase Penyebaran
Tahap ini, informasi yang telah diperoleh akan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh *user*.

3. Metode Penelitian

3.1. CRISP-DM

Suatu data analisis tidak pernah hadir dari proses sesaat, namun dari proses matang yang di guide oleh standard yang tinggi. Statement penulis di atas sangat terbuka untuk diperdebatkan namun untuk bagian standard dari pernyataan tersebut penulis yakin semua pihak akan sepakat. Dalam penelitian ini penulis menggunakan metode CRISP-DM (Cross Industry Standart Process for Data Mining) sebagai metode data mining untuk pemrosesan. Ada enam tahapan CRISP-DM yang penulis tuangkan dalam bentuk flowchart ini yaitu sebagai berikut :



Gambar 2 Flowchart CRISP-DM

Proses data mining berdasarkan CRISP-DM terdiri dari 6 fase, Yaitu:

3.1.1 Fase Pemahaman Tujuan (Business Understanding)

Pada tahapan pertama mencoba untuk memahami permasalahan yang ada dalam transaksi pembelian pada Toko Jihan. Sehingga dapat menentukan tujuan dan pola yang akan didapatkan dengan data mining. Toko Jihan dikunjungi sekitar 100 pembeli setiap harinya dengan pembelian yang berbeda-beda. Oleh karena itu pola transaksi pembelian perlu dibentuk untuk menimalisir resiko stok kosong dalam toko sehingga persediaan barang tetap terjaga dengan baik.

3.1.2 Fase Pemahaman Data (Data Understanding)

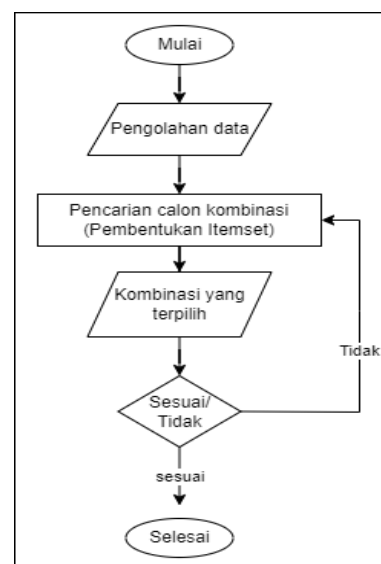
Pada tahap ini dilakukan pengumpulan data mentah atau data awal yang akan digunakan atau diolah nantinya sesuai dengan tujuan penelitian. Data yang akan digunakan adalah data dari transaksi pembelian oleh konsumen yang sudah tersimpan didalam buku yang terjadi dalam kurung waktu 3 bulan mulai dari 1 Maret 2023 sampai dengan 31 Mei 2023. Data inilah yang akan digunakan untuk analisis aturan asosiasi. Yang akan diolah untuk melihat pasangan produk yang dibeli dengan frekuensi tinggi dan menjadi pola pembelian konsumen.

3.1.3 Fase Pengolahan Data (Data Preparation)

Tahap selanjutnya data yang sudah terkumpul akan dipindahkan kedalam bentuk file excel agar memudahkan dalam proses pengolahan data dengan menghilangkan beberapa atribut yang tidak digunakan dan tidak dipakai

3.1.4 Pemodelan (Modeling)

Model yang digunakan dalam penelitian ini adalah algoritma pengambilan data dengan aturan asosiasi (Association rule) untuk menentukan hubungan asosiasi suatu kombinasi item. Cara kerja atau tahapan kerja dari Algoritma Apriori :



Gambar 3 Tahapan Algoritma Apriori

Berikut tahapan algoritma apriori:

1. Pencarian kandidat itemset (k-itemset)
2. Penetapan pola frekuensi tinggi. Kriteria item yang termasuk dalam pola frekuensi tinggi adalah item dengan support melebihi dari ketentuan minimal support
3. Jika tidak diperoleh pola frekuensi tinggi baru maka seluruh proses dihentikan.

3.1.5 Fase Evaluasi (Evaluation)

mempersiapkan dataset yang ingin diproses kemudian akan diolah menggunakan software WEKA 3.8. Tahap ini dimulai dengan mempersiapkan sampel data untuk diolah dan dianalisa. Evaluasi dilakukan disini secara rinci dan mendalam dengan tujuan untuk menyesuaikan

hasil akhir pada tahap *modeling* agar sesuai dengan target atau sasaran yang ingin diperoleh pada tahap *Business Understanding*.

4. Metode Penelitian

4.1. Proses Data Mining

4.1.1. Analisa Data Penjualan

Data yang akan digunakan adalah data transaksi pembelian barang pada Toko Jihan periode Maret sampai dengan Mei 2023 Data transaksi ini akan dilakukan proses data mining sebanyak 4 kali. Yaitu menggunakan 50, 100, 200 dan 300 data transaksi

Tabel 1 Data Transaksi

No	Tanggal Transaksi	Item
1.	11 April 2023	Mie, Gula Putih, Minyak Goreng, Sarden, Mentega
2.	11 April 2023	Mie, Gula Putih, Minyak Goreng, Susu
....		
49.	31 Maret 2023	Mie, Gula Putih, Minyak Goreng, Kopi, Susu
50.	31 Maret 2023	Minyak Goreng, Sarden, Kopi, Telur, Beras, Energen

4.1.2. Pre-processing Data

Pada tahap ini melakukan seleksi atas data diatas dan mentransformasi kedalam tabel yang dapat memudahkan proses penelitian sebagai berikut:

Tabel 2 Data Transaksi setelah di *Pre-Processing*

No	Item
1.	Mie, Gula Putih, Minyak Goreng, Sarden, Mentega
2.	Mie, Gula Putih, Minyak Goreng, Susu
....	
49.	Mie, Gula Putih, Minyak Goreng, Kopi, Susu
50.	Minyak Goreng, Sarden, Kopi, Telur, Beras, Energen

4.1.3. Menetapkan *minimum support* dan *minimum confidence*

Data transaksi penjualan dilakukan beberapa pengujian dengan menggunakan *minimum support* 20%, 50% dan *minimum confidence* 60%, 80%.

4.1.4. Permodelan menggunakan Algoritma Apriori

Sebagai *ilustrasi*, dilakukan proses pencarian aturan asosiasi dengan ketentuan nilai *minimum support* adalah 20% dan *minimum confidence* 60% menggunakan 20 Data.

Tabel 3 Tabel Data Uji

No	Item
1.	Mie, Gula Putih, Minyak Goreng, Kopi, Telur, Susu

2.	Mie, Minyak Goreng, Telur, Mentega, Beras
....	
19.	Gula Putih, Minyak Goreng, Sarden, Kopi, Telur
20.	Mie, Gula Putih, Sarden, Kopi, Telur, Susu

Iterasi-1 dilakukan untuk membentuk kandidat 1-Itemset (C1) dari data transaksi tersebut dan cari jumlah *support*-nya.. Iterasi-1 dengan *minimum support* adalah 20% dan *minimum confidence* 60% dapat dilihat dalamTtabel 4 dibawah ini:

Tabel 4 *Large-Itemset* 1

Item	Total Transaksi	Support %
Mie	15	75%
Minyak Goreng	19	95%
Kopi	16	80%
Telur	14	70%

Pada Tabel 4 terlihat *itemset* 1 yang memenuhi berjumlah 4item dengan nilai *support*-nya sama atau melebihi nilai *minimum support count* pada pengujian *minimum support* 20%, Untuk membentuk *itemset* 1 dapat menggunakan persamaan berikuu :

$$Support (Mie) = \frac{\sum \text{Transaksi item Mie}}{\sum \text{Transaksi}} \times 100\% \quad (1)$$

$$\frac{15}{20} \times 100\% = 75\%$$

Memenuhi *minimum support* 20%

Untuk mengetahui seberapa penting suatu aturan asosiasi dapat ditentukan dengan mencari *lift ratio* nya. Perhitungan nilai *lift ratio* didapat dengan menghitung *benchmark confidence* atau *expected confidence* merupakan perbandingan jumlah dari semua *item consequent* terdapat seluruh data transaksi. Yang dapat diperoleh dengan rumus berikut:

$$Bechmark confidence = \frac{Item Consequent}{Total Transaksi} \quad (2)$$

$$Lifratio = \frac{Confidence}{Benchmark Confidence} \quad (3)$$

Salah satu aturan asosisasi dapat dilihat pada tabel 5 dibawah ini:

Tabel 5 Hasil Perhitungan *Lifratio*

Itemset	Support %	Confidence%	Lifratio
Telur=> Minyak Goreng	70%	100%	1.05

Dari tabel 5 dapat diketahui nilai *lifratio* dari aturan Telur => Minyak Goreng adalah 1,05, maka dapat disimpulkan aturan tersebut dapat dikatakan valid jika mempunyai nilai *lifratio* > 1

4.2. Pengujian

Berdasarkan total pengujian yang dilakukan menggunakan 50, 100, 200, 300 data dengan nilai *minimum support* 20%, 50% dan *minimum confidence* 60% dan 80% menghasilkan aturan asosiasi yang berbeda-beda, yang dijabarkan dalam tabel 6 dibawah ini.

Tabel 6 Jumlah asosiasi berdasarkan 4 pengujian

Pengujian	50 data		100 data		200 data		300 data	
	Support	Confidence	Support	Confidence	Support	Confidence	Support	Confidence
Support	20%	50%	20%	50%	20%	50%	20%	50%
Confidence	60%	80%	60%	80%	60%	80%	60%	80%
	15	10	13	11	14	-	17	-
	7	4	8	8	4	-	-	-

Berdasarkan tabel 6 diatas ditampilkan bahwa nilai *minimum support* serta nilai *minimum confidence* terkecil adalah 20% dan 60% terlihat menghasilkan jumlah aturan asosiasi terbanyak dengan jumlah 17 aturan. Terlihat untuk pengujian dengan menggunakan *minimum support* 50% terkecil menghasilkan 4 aturan, hal ini dikarenakan jumlah jenis yang beragam dan persebaran data yang tidak merata, sehingga banyak item yang tidak memenuhi jumlah *minimum confidence*.

Dari empat pengujian yang dilakukan, pengujian ketiga dengan 200 data transaksi menggunakan *minimum support* 20% dan *minimum confidence* 80% merupakan kombinasi aturan yang bisa direkomendasi oleh peneliti untuk menentukan barang yang sering dibeli oleh konsumen terdapat pada tabel 7 dibawah ini :

Tabel 7 Hasil asosiasi *support* 20% dan *confidence* 80%

Aturan	Lifratio	Support	Confidence
Minyak Goreng, Kopi => Telur	1.138	49%	87%
Telur, Susu => Minyak Goreng	1.013	47%	89%
Susu, Kopi=> Minyak Goreng	1.051	39%	92%

Susu, Kopi => Telur	1.146	37%	88%
---------------------	-------	-----	-----

Dilihat dari tabel diatas, hasil dari aturan asosiasi yang terbentuk untuk setiap aturan asosiasi mempunyai nilai *lifratio* > 1. Dapat disimpulkan bahwa aturan asosiasi atau *association rule* yang dihasilkan *valid* dan dapat dijadikan untuk membuat strategi pemasaran.

5. Penutup

Berdasarkan hasil penelitian peneliti berhasil menganalisis data transaksi pembelian Toko Jihan dari rentang waktu bulan Maret 2023 – Mei 2023 dengan jumlah transaksi sebanyak 300 data transaksi pembelian. Algoritma apriori mampu digunakan untuk menentukan barang yang paling sering dibeli konsumen dengan melihat kecenderungan konsumen dalam melakukan transaksi. Hasil analisis yang didapatkan setelah menggunakan perbandingan *minimum support* dan *minimum confidence* yang berbeda-beda berdasarkan data transaksi yang ada adalah dengan menggunakan *minimum support* 20% (kuatnya kombinasi *item* tersebut dalam *database*) dan *minimum confidence* 80% (kuatnya hubungan antar *item* dalam aturan asosiasi) menghasilkan empat aturan asosiasi. Salah satu contohnya yaitu jika konsumen membeli Minyak goreng, Kopi maka 87% (kepastian kosumen dalam membeli *item*) akan membeli Telur. Dari data hasil aturan yang telah diperoleh, dapat diketahui barang apa saja yang sering dibeli secara bersamaan oleh setiap konsumen. Informasi ini sangat berguna, dengan hal tersebut pihak toko dapat membuat keputusan bisnis dengan mengatur tata letak barang.

Daftar Pustaka

- [1] R. Rachmawati, “Peranan Bauran Pemasaran (Marketing Mix) terhadap Peningkatan Penjualan,” *J. Kompetensi Tek.*, vol. 2, no. 2, pp. 143–150, 2011.
- [2] Y. Ardilla et al., DATA MINING DAN APLIKASINYA. Penerbit Widina, 2021. [Online]. Available: <https://books.google.co.id/books?id=53FXEA AAQBAJ>
- [3] R. F. Putra et al., DATA MINING : Algoritma dan Penerapannya. PT. Sonpedia Publishing Indonesia, 2023. [Online]. Available: <https://books.google.co.id/books?id=zLHGE AAAQBAJ>
- [4] E. Buulolo, Data Mining Untuk Perguruan Tinggi. Deepublish, 2020. [Online]. Available: https://books.google.co.id/books?id=-K_SDwAAQBAJ
- [5] S. T. M. T. I. Bernardus Ari Kuncoro, Lima Dasar Data Science untuk Pemula. Pustaka Media Guru, 2023. [Online]. Available:

<https://books.google.co.id/books?id=xRusEAAQBAJ>

- [6] I. Budiman, T. Prahasto, and Y. Christyono, "Data Clustering Menggunakan Metodologi CRISP-DM

Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma," *J. Sist. Inf. Bisnis*, vol. 1, no. 3, pp. 15–16, 2014, doi: 10.21456/vol1iss3pp129-134.