



Implementasi Data Mining Asosiasi Terhadap Data Penjualan

Mohammad Khoirul Anwar¹, Indra Wahyu Mahendra², Chirza Rahman³, Muhammad Arifin⁴

^{1,2,3,4}Fakultas Teknik, Sistem Informasi, Universitas Muria Kudus, Kota, Indonesia

Email: 1*khoirulawr754@gmail.com, 2*indramahendra13@gmail.com, 3*chirzarhman@gmail.com,

4*arifin.m@umk.ac.id,

(* : coresponding author)

Abstrak— Penelitian ini membahas penerapan algoritma Apriori dalam data mining untuk menemukan pola asosiasi dalam data transaksi penjualan produk. Tujuan utama dari penelitian ini adalah mengidentifikasi produk-produk yang sering dibeli bersama, sehingga dapat memberikan wawasan untuk pengoptimalan strategi penjualan dan penempatan produk. Proses ini dilakukan dengan menggunakan data transaksi dari sebuah platform e-commerce, yang dibagi berdasarkan kelompok usia dan jenis kelamin pelanggan. Hasil dari analisis menunjukkan bahwa terdapat perbedaan pola pembelian produk berdasarkan faktor demografis seperti usia dan jenis kelamin, yang memberikan informasi penting dalam merancang strategi pemasaran yang lebih tepat sasaran. Analisis ini dilakukan dengan menggunakan Google Colab, yang menghasilkan nilai support, confidence, dan lift yang signifikan dalam mengukur kekuatan hubungan antara produk-produk yang dibeli bersamaan. Temuan ini diharapkan dapat memberikan kontribusi pada pengambilan keputusan dalam pengelolaan produk dan promosi yang lebih efektif di sektor ritel.

Kata Kunci: Data Mining, Algoritma Apriori, Aturan Asosiasi, Data Transaksi, Optimasi Penjualan

Abstract— *This research discusses the application of the Apriori algorithm in data mining to find association patterns in product sales transaction data. The main objective of this research is to identify products that are often purchased together, so as to provide insights for optimizing sales strategies and product placement. This process was carried out using transaction data from an e-commerce platform, which was divided by age group and gender of customers. The results of the analysis show that there are differences in product purchasing patterns based on demographic factors such as age and gender, which provide important information in designing more targeted marketing strategies. This analysis was conducted using Google Colab, which produces significant support, confidence, and lift values in measuring the strength of the relationship between products purchased together. The findings are expected to contribute to decision-making in product management and more effective promotions in the retail sector.*

Keywords: *Data Mining, Apriori Algorithm, Association Rules, Transaction Data, Sales Optimization*

1. PENDAHULUAN

Di Indonesia perkembangan dunia bisnis sudah sangat pesat, khususnya dalam bidang penjualan jasa maupun ritel, persaingan dunia bisnis menuntut para pelaku untuk selalu berkembang dalam strategi pemasaran mereka. Salah satu tantangan utama yang harus dihadapi adalah cara pengelolaan dan pemanfaatan data transaksi penjualan yang selalu meningkat secara signifikan. Data transaksi dalam jumlah besar mempunyai potensi informasi yang tersembunyi dan dapat digunakan untuk pengambilan keputusan strategis, seperti penentuan paket promosi, pengaturan stok barang, hingga rekomendasi produk untuk pelanggan.

Dalam dunia e-commerce dan ritel, mengetahui produk mana yang sering dibeli bersama merupakan informasi yang sangat berharga untuk merancang strategi pemasaran dan penempatan produk yang lebih baik(Kristania & Listanto, 2022). Salah satu teknik dalam data mining yang sering digunakan untuk menggali informasi ini adalah algoritma Apriori. Algoritma ini bekerja dengan cara menemukan asosiasi atau pola yang sering muncul bersama dalam dataset transaksi.

Penelitian terkait dalam bidang algoritma Apriori telah banyak dilakukan, namun implementasi secara spesifik pada data transaksi penjualan dari platform e-commerce yang mempertimbangkan faktor demografis pelanggan, seperti usia dan jenis kelamin, masih cukup terbatas. Oleh karena itu, penelitian ini berfokus pada penerapan algoritma Apriori untuk menganalisis data transaksi yang mencakup kelompok usia dan jenis kelamin pelanggan, dengan tujuan untuk menemukan asosiasi yang dapat dimanfaatkan untuk strategi pemasaran yang lebih

terarah. Persaingan yang tinggi dalam bisnis ritel dan kuliner mengharuskan pemilik usaha memanfaatkan data transaksi untuk mendapatkan strategi pemasaran yang lebih presisi(Syahputri, 2020; Takdirillah, 2020)

2. METODE

2.1 Data Mining

Data mining merupakan proses mengekstraksi informasi, pengetahuan, atau pola yang bermakna dari kumpulan data yang besar dengan menggunakan pembelajaran mesin (machine learning), teknik statistic, sistem manajemen basis data, dan matematika. Tahap ini bertujuan untuk mengidentifikasi pola-pola yang tidak tampak secara langsung. Menurut Saputra dan Sibarani(Sibarani, 2020), data mining menggabungkan teknik dari berbagai disiplin ilmu seperti statistik, pembelajaran mesin, dan pengenalan pola untuk mengekstraksi informasi yang berguna dari database besar. Proses ini juga disebut dengan istilah knowledge discovery in database (KDD)(Takdirillah, 2020), mencakup sejumlah langkah mulai dari seleksi data, transformasi, pra-pemrosesan, dan interpretasi hasilnya, dan penerapan teknik penggalian(Erwansyah et al., 2021).

2.2 Tahapan Penelitian

Data yang digunakan dalam penelitian ini diperoleh dari platform e-commerce yang berisi informasi transaksi penjualan(Sahara et al., 2022). Data ini mencakup informasi produk yang dibeli, kelompok usia dan jenis kelamin pelanggan, serta tanggal transaksi. Sebelum diterapkan algoritma Apriori, data terlebih dahulu dibersihkan dengan cara menangani missing values dan pencihan. Proses pembersihan ini bertujuan untuk meningkatkan kualitas data yang digunakan dalam analisis. Penelitian lain juga menggunakan pendekatan sistematis seperti SDLC(Muchlis et al., 2021) dan CRISP-DM(Arinal & Melani, 2023) dalam mengembangkan solusi data mining.

2.2. Penerapan Algoritma Apriori

Apriori adalah salah satu algoritma paling umum dalam asosiasi karena kesederhanaan dan efektivitasnya dalam menemukan pola frekuensi tinggi antar item(Erwansyah et al., 2021; Sibarani, 2020). Algoritma Apriori diterapkan untuk menemukan asosiasi atau aturan antara produk yang sering dibeli bersama dalam transaksi. Aturan asosiasi yang ditemukan dihitung menggunakan tiga metrik utama:

- a. **Support:** Menunjukkan seberapa sering suatu itemset (kombinasi produk) muncul dalam dataset.
- b. **Confidence:** Mengukur seberapa besar kemungkinan produk kedua dibeli setelah produk pertama dibeli.
- c. **Lift:** Menunjukkan kekuatan hubungan antara dua produk, dengan nilai lift yang lebih tinggi menunjukkan hubungan yang lebih kuat dari yang diharapkan secara acak.

Implementasi algoritma Apriori dilakukan menggunakan bahasa pemrograman Python di platform Google Colab. Dataset dibagi berdasarkan kelompok usia (18-25, 26-35, 46+) dan jenis kelamin (Pria, Wanita) untuk menganalisis pola pembelian yang berbeda di antara pelanggan.

3. ANALISA DAN PEMBAHASAN

Hasil dari penerapan algoritma Apriori menunjukkan beberapa asosiasi menarik yang dapat membantu memahami pola pembelian pelanggan berdasarkan demografi mereka. Berikut adalah beberapa aturan asosiasi yang ditemukan:

3.1 Hasil Data Mining Aturan Asosiasi

- a. **Aturan 1 :** Kelompok Usia 26-35 (Pelanggan Wanita) → Produk Kecantikan
 1. **Support:** 10,2%
 2. **Confidence:** 50,2%
 3. **Lift:** 1,06

Hasil ini menunjukkan bahwa lebih dari 50% pelanggan wanita usia 26-35 yang membeli produk tertentu juga membeli produk kecantikan, dengan hubungan yang cukup kuat berdasarkan nilai lift.

- b. **Aturan 2** : Kategori Produk: Kecantikan → Pelanggan Wanita

4. **Support:** 10,2%
5. **Confidence:** 54,1%
6. **Lift:** 1,06

Aturan ini menunjukkan bahwa sekitar 54% pelanggan yang membeli produk kecantikan adalah wanita, dengan hubungan positif yang kuat antara kedua variabel ini.

- c. **Aturan 3** : Kelompok Usia 46+ → Produk Wanita

7. **Support:** 22,4%
8. **Confidence:** 50,8%
9. **Lift:** 0,99

Meskipun nilai lift lebih rendah dibandingkan dengan aturan sebelumnya, hasil ini menunjukkan bahwa pelanggan usia 46+ cenderung membeli produk untuk wanita, yang menunjukkan bahwa mereka mungkin membeli untuk anggota keluarga mereka.

- d. **Aturan 4** : Kelompok Usia 18-25 (Pelanggan Pria) → Produk Pria

10. **Support:** 7,7%
11. **Confidence:** 51,6%
12. **Lift:** 1,05

Hasil ini menunjukkan bahwa ada hubungan positif antara kelompok usia muda pria dengan pembelian produk pria, dengan tingkat kepercayaan yang lebih tinggi dibandingkan dengan aturan lainnya.

3.2 Visualisasi dan Wawasan

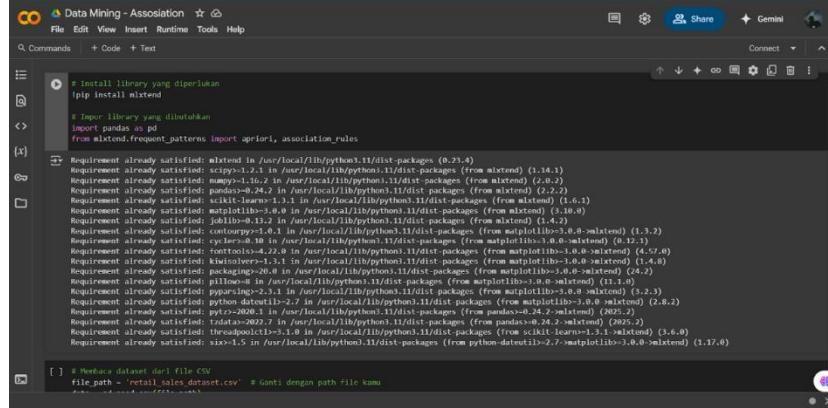
Untuk memperjelas hubungan antara support, confidence, dan lift, kami menggunakan beberapa visualisasi:

- e. **Scatter Plot** : Menunjukkan hubungan antara support dan confidence dari setiap aturan asosiasi. Plot ini memberikan gambaran visual mengenai seberapa kuat hubungan antara produk yang dibeli bersama.
- f. **Bar Chart**: Menampilkan 10 aturan asosiasi teratas berdasarkan nilai lift. Ini membantu mengidentifikasi produk yang paling sering dibeli bersama dengan kekuatan asosiasi yang lebih besar.
- g. **Pie Chart**: Menunjukkan distribusi support antar kategori produk, memberikan wawasan tentang kategori mana yang lebih sering dibeli bersama.

Visualisasi ini membantu dalam pengambilan keputusan terkait strategi pemasaran, penempatan produk, dan promosi.

3.3 Modeling

Dalam tahap pemodelan, peneliti melakukan pengujian data melalui platform Google Colab dengan memanfaatkan metode asosiasi, khususnya algoritma Apriori, guna menganalisis pola pembelian. Implementasi antarmuka yang dihasilkan menggunakan Google Colab dapat diamati seperti tampilan berikut:



```
# Install library yang diperlukan
!pip install mxnet

# Impor library yang dibutuhkan
import pandas as pd
from mxnet.frequent_patterns import apriori, association_rules

# Membaca dataset dari file CSV
file_path = 'retail_sales_dataset.csv' # Ganti dengan path file kamu
data = pd.read_csv(file_path)

# Melihat beberapa baris pertama dari dataset
data.head()
```

Gambar 1.1 Tampilan Google Colab

Hasil pengujian yang dilakukan melalui platform Google Colab dapat dilihat pada tampilan berikut:

a. Pre Coding

Beberapa library python yang digunakan, adalah:



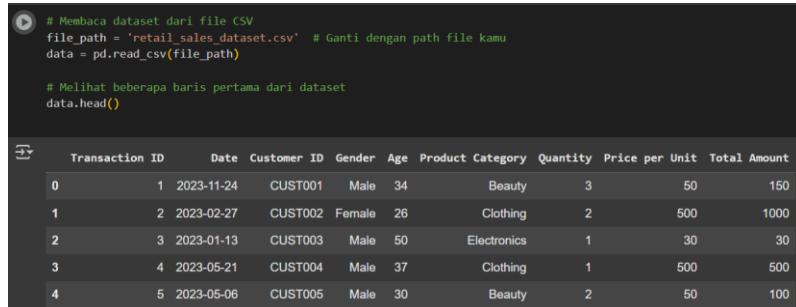
```
# Install library yang diperlukan
!pip install mxnet

# Impor library yang dibutuhkan
import pandas as pd
from mxnet.frequent_patterns import apriori, association_rules
```

Gambar 1.2 Install library python

b. Read and show dataset

Gambar ini menampilkan 5 baris pertama dataset penjualan retail yang dibaca dari file CSV menggunakan pandas. Dataset berisi kolom-kolom transaksi seperti ID pelanggan, produk, jumlah, dan harga, contohnya transaksi produk Beauty 3 unit seharga 150 oleh pelanggan pria 34 tahun.

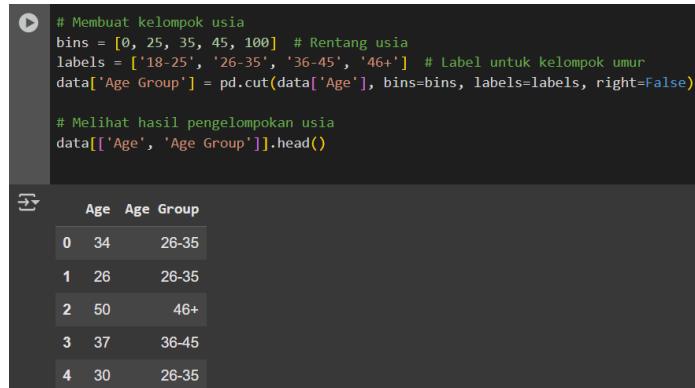


	Transaction ID	Date	Customer ID	Gender	Age	Product	Category	Quantity	Price per Unit	Total Amount
0	1	2023-11-24	CUST001	Male	34	Beauty		3	50	150
1	2	2023-02-27	CUST002	Female	26	Clothing		2	500	1000
2	3	2023-01-13	CUST003	Male	50	Electronics		1	30	30
3	4	2023-05-21	CUST004	Male	37	Clothing		1	500	500
4	5	2023-05-06	CUST005	Male	30	Beauty		2	50	100

Gambar 1.3 Baca dan menampilkan file CSV

c. Pengelompokan dan menampilkan hasil

Gambar ini menunjukkan pengelompokan usia pelanggan dalam dataset penjualan retail menggunakan pandas. Kode pd.cut() membagi usia menjadi 4 kategori (18-25, 26-35, 36-45, 46+), seperti contoh pelanggan usia 34 tahun masuk kelompok 26-35 dan usia 50 tahun masuk kategori 46+. Hasilnya ditampilkan dalam kolom baru 'Age Group' bersanding dengan usia asli.



```
# Membuat kelompok usia
bins = [0, 25, 35, 45, 100] # Rentang usia
labels = ['18-25', '26-35', '36-45', '46+'] # Label untuk kelompok umur
data['Age Group'] = pd.cut(data['Age'], bins=bins, labels=labels, right=False)

# Melihat hasil pengelompokan usia
data[['Age', 'Age Group']].head()
```

	Age	Age Group
0	34	26-35
1	26	26-35
2	50	46+
3	37	36-45
4	30	26-35

Gambar 1.4 Pengelompokan pembelian

d. Preprocessing data analisis asosiasi

Gambar ini menunjukkan proses persiapan data untuk analisis asosiasi menggunakan one-hot encoding. Data kolom Gender, Age Group, dan Product Category diubah menjadi format biner (True/False), dimana setiap kategori diubah menjadi kolom terpisah. Contohnya, pelanggan pertama (baris 0) adalah Male (True), kelompok usia 26-35 (True), dan membeli produk Beauty (True).



```
[1] # Membaca data untuk analisis asosiasi
transaction_data = data[['Gender', 'Age Group', 'Product Category']]

# Menggunakan One-Hot Encoding untuk kolom Gender, Age Group, dan Product Category
encoded_data = pd.get_dummies(transaction_data, columns=['Gender', 'Age Group', 'Product Category'])

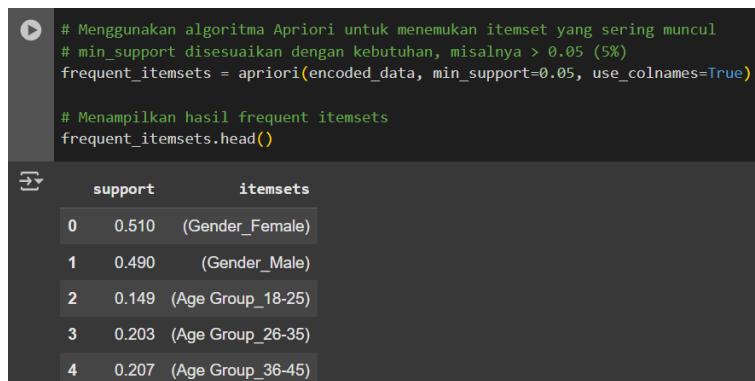
# Melihat data yang sudah terencode
encoded_data.head()
```

	Gender_Female	Gender_Male	Age Group_18-25	Age Group_26-35	Age Group_36-45	Age Group_46+	Product Category_Beauty	Product Category_Clothing	Product Category_Electronics
0	False	True	False	True	False	False	True	False	False
1	True	False	False	True	False	False	False	True	False
2	False	True	False	False	False	True	False	False	True
3	False	True	False	False	True	False	False	True	False
4	False	True	False	True	False	False	True	False	False

Gambar 1.5 Preprocessing data

e. Analisis Pola Belanja Apriori

Gambar ini menampilkan hasil analisis pola belanja menggunakan algoritma Apriori dengan minimum support 5%. Tabel menunjukkan itemset yang sering muncul beserta nilai support-nya, seperti Gender_Female (51%), Gender_Male (49%), dan kelompok usia 26-35 tahun (20.3%).



```
# Menggunakan algoritma Apriori untuk menemukan itemset yang sering muncul
# min_support disesuaikan dengan kebutuhan, misalnya > 0.05 (5%)
frequent_itemsets = apriori(encoded_data, min_support=0.05, use_colnames=True)

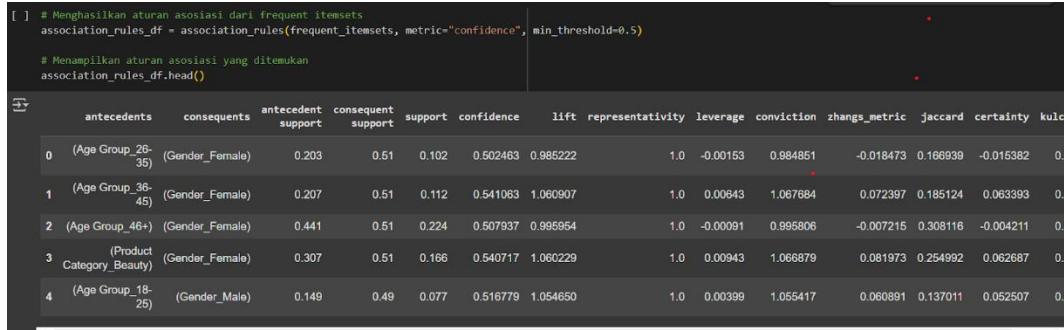
# Menampilkan hasil frequent itemsets
frequent_itemsets.head()
```

	support	itemsets
0	0.510	(Gender_Female)
1	0.490	(Gender_Male)
2	0.149	(Age Group_18-25)
3	0.203	(Age Group_26-35)
4	0.207	(Age Group_36-45)

Gambar 1.6 Analisis pola

f. Aturan Asosiasi Pola Belanja

Gambar ini menampilkan hasil analisis asosiasi belanja menggunakan algoritma Apriori, mengungkap pola seperti: pembeli perempuan mendominasi kelompok usia 26-35 tahun (50.2%) dan pembelian produk kecantikan (54.1%), sementara pembeli laki-laki lebih banyak di usia 18-25 tahun (51.7%).



The screenshot shows a Jupyter Notebook cell with the following code:

```
[ ] # Menghasilkan aturan asosiasi dari frequent itemsets
association_rules_df = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.5)

# Menampilkan aturan asosiasi yang ditemukan
association_rules_df.head()
```

The resulting DataFrame is displayed:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage	conviction	zhangs_metric	jaccard	certainty	kulcs
0	(Age Group_26-35)	(Gender_Female)	0.203	0.51	0.102	0.502463	0.985222		1.0	-0.00153	0.984851	-0.018473	0.166939	-0.015382
1	(Age Group_36-45)	(Gender_Female)	0.207	0.51	0.112	0.541083	1.060907		1.0	0.00643	1.067884	0.072397	0.185124	0.063393
2	(Age Group_46+)	(Gender_Female)	0.441	0.51	0.224	0.507937	0.995984		1.0	-0.00091	0.985806	-0.007215	0.308116	-0.004211
3	(Product Category_Beauty)	(Gender_Female)	0.307	0.51	0.166	0.540717	1.060229		1.0	0.00943	1.066879	0.081973	0.254992	0.062687
4	(Age Group_18-25)	(Gender_Male)	0.149	0.49	0.077	0.516779	1.054650		1.0	0.00399	1.055417	0.060891	0.137011	0.052507

Gambar 1.7 Aturan Asosiasi

g. Ekspor Hasil Analisis Apriori

Gambar ini menunjukkan kode Python untuk menyimpan hasil analisis aturan asosiasi ke file CSV (hasil_apriori.csv) dan mengunduhnya langsung dari Google Colab, memungkinkan penggunaan hasil analisis pola belanja untuk kebutuhan lebih lanjut.



```
[ ] # prompt: buatkan code dari hasil apriori tersebut, dan saya tinggal download file csv nya

import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules

# ... (previous code from the prompt)

# Save the association rules to a CSV file
association_rules_df.to_csv('hasil_apriori.csv', index=False)

# Download the CSV file
from google.colab import files
files.download('hasil_apriori.csv')
```

Gambar 1.8 Ekspor Hasil

h. Visualisasi Aturan Asosiasi

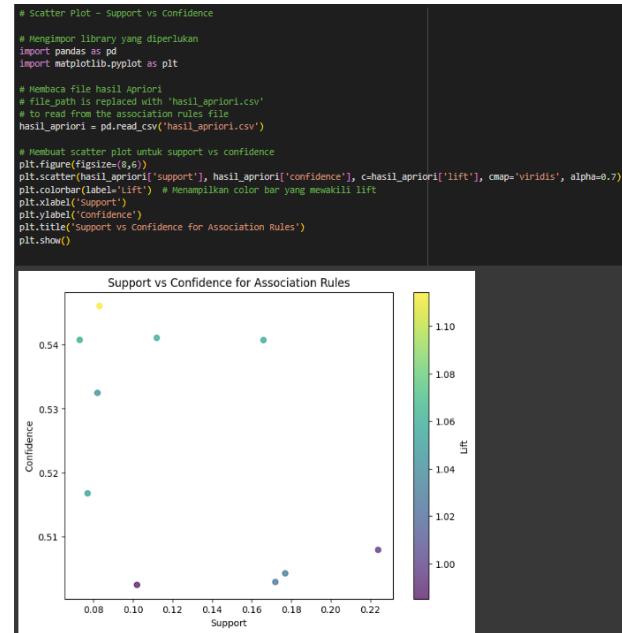
Gambar ini menampilkan visualisasi hubungan antara Support dan Confidence dari aturan asosiasi dalam bentuk scatter plot, dengan warna titik menunjukkan nilai Lift, dimana semakin terang warna menunjukkan kekuatan hubungan yang lebih kuat antara item.

JRIIN : Jurnal Riset Informatika dan Inovasi

Volume 2, No. 12 Mei 2025

ISSN 3025-0919 (media online)

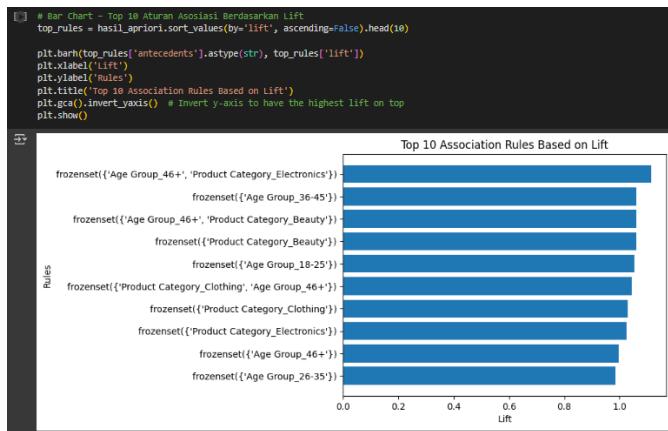
Hal 2124 - 2132



Gambar 1.9 Visualisasi Aturan Asosiasi

i. 10 Aturan Asosiasi

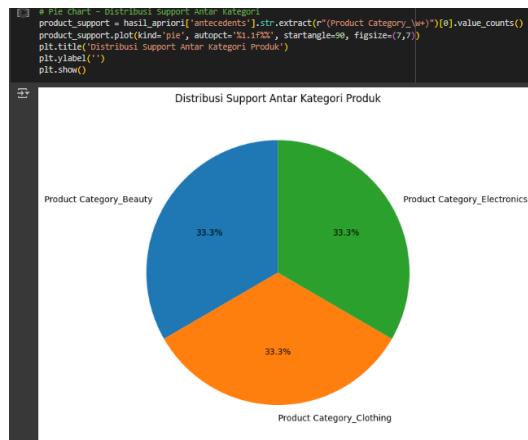
Gambar ini menampilkan 10 aturan asosiasi terkuat berdasarkan nilai lift dalam bentuk diagram batang, di mana kombinasi "Kelompok Usia 46+ & Produk Elektronik" menempati posisi teratas, menunjukkan hubungan paling signifikan antara variabel-variabel tersebut.



Gambar 1.9 Aturan Asosiasi

j. Distribusi Kategori Produk

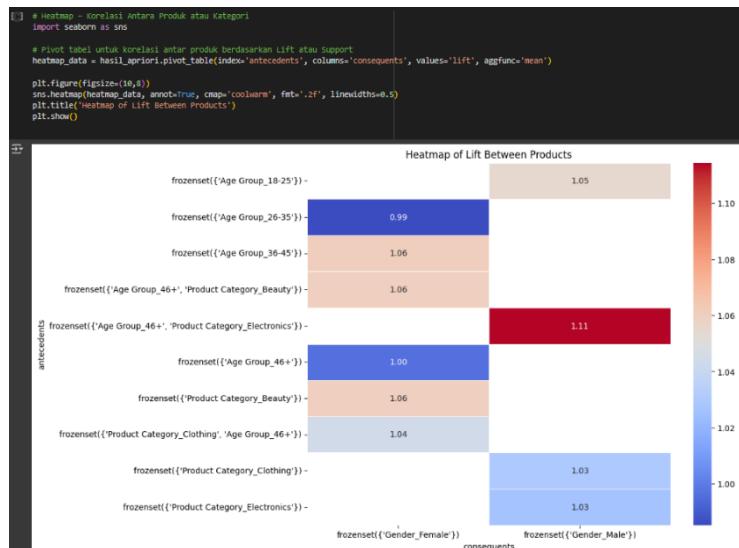
Gambar ini menampilkan distribusi support antar kategori produk dalam bentuk diagram pie, menunjukkan persentase keterkaitan masing-masing kategori produk (Beauty, Electronics, Clothing) dalam aturan asosiasi yang dihasilkan.



Gambar 1.10 Distribusi Kategori Produk

k. Heatmap Korelasi Produk

Gambar ini menampilkan heatmap yang memvisualisasikan kekuatan hubungan (lift) antar produk/kategori dalam aturan asosiasi, dengan warna hangat menunjukkan korelasi kuat dan warna dingin menunjukkan korelasi lemah, memudahkan identifikasi pola hubungan antar item.



Gambar 1.11 Heatmap Korelasi Produk

l. Distribusi Nilai Confidence

Gambar ini menampilkan distribusi nilai confidence dari seluruh aturan asosiasi dalam bentuk histogram, menunjukkan bahwa sebagian besar aturan memiliki confidence berkisar antara 0.51 hingga 0.54.



Gambar 1.12 Distribusi Nilai Confidence

4. KESIMPULAN

Dari hasil analisis yang dilakukan, dapat disimpulkan bahwa algoritma Apriori dapat digunakan untuk mengidentifikasi asosiasi yang kuat antara produk yang sering dibeli bersama, berdasarkan kelompok usia dan jenis kelamin pelanggan. Hasil penelitian menunjukkan bahwa wanita usia 26-35 lebih sering membeli produk kecantikan, sedangkan pelanggan usia lebih tua (46+) cenderung membeli produk untuk wanita. Wawasan ini dapat digunakan untuk meningkatkan penempatan produk di toko dan mengarahkan promosi yang lebih tepat sesuai dengan kelompok pelanggan yang relevan.

Untuk penelitian selanjutnya, bisa dilakukan perpanjangan analisis dengan memasukkan lebih banyak faktor demografis atau menggunakan teknik data mining yang lebih canggih untuk meningkatkan akurasi prediksi dan pemahaman pola pembelian.

REFERENCES

- Arinal, V., & Melani, M. A. (2023). Penerapan Metode Asosiasi Pada Data Penjualan Transaksi Menggunakan Algoritma Apriori (Studi Kasus Circle'K Apartemen Marabella Jakarta Selatan). *Jurnal Sains Dan Teknologi*, 5(1), 170–176.
- Erwansyah, K., Andika, B., & Gunawan, R. (2021). Implementasi Data Mining Menggunakan Asosiasi Dengan Algoritma Apriori Untuk Mendapatkan Pola Rekomendasi Belanja Produk Pada Toko Avis Mobile. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, 4(1), 148. <https://doi.org/10.53513/jsk.v4i1.2628>
- Kristania, Y. M., & Listanto, S. (2022). Implementasi Data Mining Terhadap Data Penjualan Dengan Algoritma Apriori Pada Pt. Duta Kencana Swaguna. *Jurnal Teknoinfo*, 16(2), 364. <https://doi.org/10.33365/jti.v16i2.1973>
- Muchlis, M. M., Fitri, I., & Nuraini, R. (2021). Rancang Bangun Aplikasi Data Mining pada Penjualan Distro Bloods Berbasis Web menggunakan Algoritma Apriori. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 4(2), 26. <https://doi.org/10.35870/jtik.v5i1.197>
- Sahara, W., Saragih, S. D., & Windarto, A. P. (2022). Teknik Asosiasi Datamining Dalam Menentukan Pola Penjualan dengan Metode Apriori. *TIN: Terapan Informatika Nusantara*, 2(12), 684–689. <https://doi.org/10.47065/tin.v2i12.1577>
- Sibarani, A. J. P. (2020). Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(2), 262–276. <https://doi.org/10.35957/jatisi.v7i2.195>
- Syahputri, N. (2020). Penerapan Data Mining Asosiasi pada Pola Transaksi dengan Metode Apriori. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 4(2), 728–736.
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan. *Edumatic : Jurnal Pendidikan Informatika*, 4(1), 37–46. <https://doi.org/10.29408/edumatic.v4i1.2081>