

ASSOCIATION RULE MINING

1. Introduction

Association Rule Mining merupakan teknik dalam data mining yang digunakan untuk menemukan pola hubungan antar item dalam suatu dataset. Teknik ini banyak digunakan dalam analisis transaksi, seperti market basket analysis, untuk mengetahui item apa saja yang sering muncul bersamaan.

Sebagai contoh, dalam konteks e-commerce, kita dapat mengetahui bahwa pelanggan yang membeli produk A cenderung juga membeli produk B. Informasi ini sangat berguna untuk:

- strategi promosi (bundling)
- rekomendasi produk
- penataan produk

2. Support for an Itemset

Support digunakan untuk mengukur seberapa sering suatu itemset muncul dalam dataset.

Rumus Support

$$\text{Support}(A) = (\text{jumlah transaksi yang mengandung } A) / (\text{total transaksi})$$

Untuk dua item:

$$\text{Support}(A \cup B) = (\text{jumlah transaksi AdanB}) / (\text{total transaksi})$$

3. Association Rules

Association rule berbentuk:

$$A \rightarrow B$$

Artinya:

Jika A terjadi, maka B cenderung terjadi.

Confidence

Mengukur seberapa besar kemungkinan B terjadi jika A terjadi.

$$\text{Confidence}(A \rightarrow B) = \text{Support}(A \cup B) / \text{Support}(A)$$

Lift

Mengukur kekuatan hubungan A dan B.

$$\text{Lift}(A \rightarrow B) = \text{Confidence}(A \rightarrow B) / \text{Support}(B)$$

Interpretasi:

- Lift > 1 → hubungan kuat
- Lift = 1 → independen
- Lift < 1 → hubungan lemah

4. Finding the Best N Rules

Tidak semua aturan menarik. Oleh karena itu, kita memilih **N aturan terbaik** berdasarkan:

- Confidence tinggi
- Lift tinggi
- Support cukup besar

Tujuannya:

- menghindari aturan yang tidak relevan
- fokus pada pola yang benar-benar berguna

5. Generating Association Rules

Langkah umum:

1. Hitung support semua itemset
2. Pilih itemset yang memenuhi minimum support
3. Bentuk aturan ($A \rightarrow B$)
4. Hitung confidence dan lift
5. Filter aturan terbaik

6. Apriori Algorithm

Apriori adalah algoritma untuk menemukan frequent itemset dengan prinsip:

Jika suatu itemset tidak frequent, maka semua superset-nya juga tidak akan frequent

Proses Apriori

1. Tentukan minimum support
2. Cari frequent 1-itemset
3. Gabungkan menjadi 2-itemset
4. Ulangi sampai tidak ada kombinasi baru

STUDI KASUS

Sebuah perusahaan e-commerce ingin menganalisis pola pembelian pelanggan berdasarkan data transaksi.

Data contoh (transaksi):

- T1: Roti, Susu
- T2: Roti, Popok, Bir, Telur
- T3: Susu, Popok, Bir, Cola
- T4: Roti, Susu, Popok, Bir
- T5: Roti, Susu, Popok, Cola

Yang Harus Dianalisis

1. Hitung support setiap item
2. Tentukan frequent itemset (min support)
3. Bentuk aturan asosiasi
4. Hitung confidence
5. Hitung lift
6. Tentukan aturan terbaik (Top-N rules)
7. Interpretasi pola pembelian

LANGKAH PENYELESAIAN

LANGKAH 1: IMPORT LIBRARY

Penjelasan:

Digunakan library untuk mengolah data transaksi dan menghasilkan association rules.

```
import pandas as pd # manipulasi data
from mlxtend.preprocessing import TransactionEncoder # encoding transaksi
from mlxtend.frequent_patterns import apriori, association_rules # apriori
```

LANGKAH 2: INPUT DATA TRANSAKSI

Penjelasan:

Data transaksi disimpan dalam bentuk list, di mana setiap baris mewakili satu transaksi.

```
transactions = [
    ['Roti', 'Susu'],
    ['Roti', 'Popok', 'Bir', 'Telur'],
    ['Susu', 'Popok', 'Bir', 'Cola'],
    ['Roti', 'Susu', 'Popok', 'Bir'],
    ['Roti', 'Susu', 'Popok', 'Cola']
]
```

LANGKAH 3: TRANSFORMASI DATA

Penjelasan:

Data transaksi diubah menjadi bentuk matriks biner agar bisa diproses oleh algoritma Apriori.

```
te = TransactionEncoder()
te_data = te.fit(transactions).transform(transactions)

df = pd.DataFrame(te_data, columns=te.columns_)
print(df)
```

LANGKAH 4: MENCARI FREQUENT ITEMSET

Penjelasan:

Menggunakan algoritma Apriori untuk mencari kombinasi item yang sering muncul.

```
frequent_itemsets = apriori(df, min_support=0.4, use_colnames=True)
print(frequent_itemsets)
```

LANGKAH 5: GENERATE ASSOCIATION RULES

Penjelasan:

Membentuk aturan asosiasi dari frequent itemset dan menghitung metrik seperti confidence dan lift.

```
rules = association_rules(frequent_itemsets, metric="confidence",
min_threshold=0.6)

print(rules)
```

LANGKAH 6: MENENTUKAN TOP-N RULES

Penjelasan:

Mengambil aturan terbaik berdasarkan nilai lift tertinggi.

```
top_rules = rules.sort_values(by='lift', ascending=False).head(5)

print(top_rules)
```

LANGKAH 7: VISUALISASI NETWORK GRAPH

Network graph digunakan untuk menunjukkan hubungan antar item:

- Node = produk
- Edge = aturan asosiasi
- Arah panah = arah rule ($A \rightarrow B$)

Visualisasi ini sangat membantu untuk:

- melihat pola pembelian
 - menemukan item yang sering menjadi pusat hubungan
-

```
import networkx as nx
import matplotlib.pyplot as plt

# buat graph
G = nx.DiGraph()

# tambahkan edge dari rules
for _, row in rules.iterrows():
    for ant in row['antecedents']:
        for cons in row['consequents']:
            G.add_edge(ant, cons, weight=row['lift'])

# gambar graph
```

```
plt.figure()

pos = nx.spring_layout(G) # posisi node otomatis

# gambar node
nx.draw_networkx_nodes(G, pos)

# gambar edge
nx.draw_networkx_edges(G, pos, arrows=True)

# label node
nx.draw_networkx_labels(G, pos)

plt.title("Network Graph Association Rules")
plt.show()
```

INTERPRETASI VISUALISASI

Network Graph

- Node yang banyak terhubung → produk populer
 - Panah menunjukkan arah hubungan
 - Jika banyak panah menuju satu item → item tersebut sering dibeli bersama item lain
-

LANGKAH 9: INTERPRETASI HASIL

Penjelasan:

Interpretasi dilakukan dengan melihat aturan yang terbentuk. Jelaskan interpretasi dari hasil yang diperoleh !

KESIMPULAN

Association Rule Mining memungkinkan kita menemukan pola tersembunyi dalam data transaksi. Dengan menggunakan algoritma Apriori, kita dapat mengidentifikasi kombinasi item yang sering muncul dan membentuk aturan asosiasi yang berguna untuk pengambilan keputusan bisnis. Pemilihan aturan terbaik berdasarkan support, confidence, dan lift sangat penting agar hasil analisis relevan dan tidak menyesatkan. Oleh karena itu, teknik ini sangat bermanfaat dalam sistem rekomendasi, strategi pemasaran, dan analisis perilaku pelanggan.