



DATA MINING

PERTEMUAN 8 (*Association Rule Mining*)

Memahami dan menjelaskan *Mining Associations rule* dan menerapkan pada data





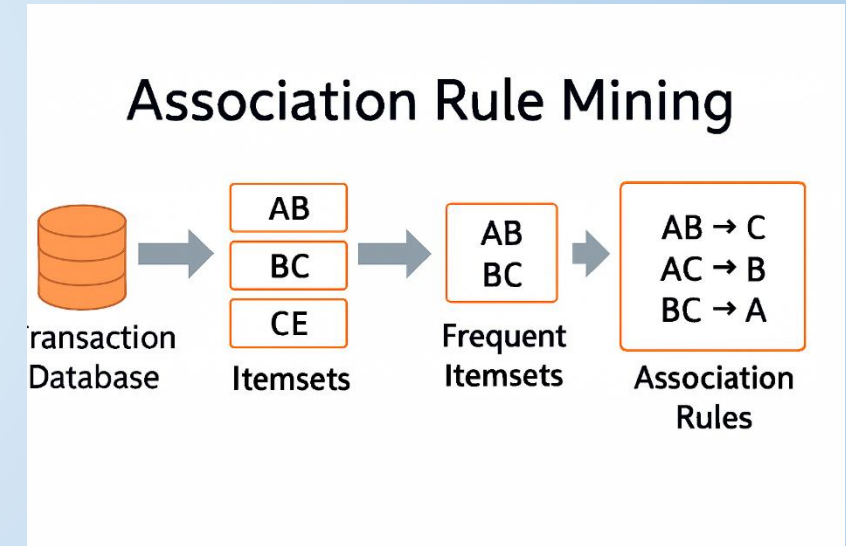
DATA MINING

Apa itu Asosiasi?

Association Rule Mining adalah sebuah teknik dalam data mining untuk menemukan **hubungan atau asosiasi** yang menarik antar item dalam sebuah dataset berukuran besar.

Tujuan utamanya adalah untuk mengidentifikasi aturan-aturan kuat dalam format "**Jika A terjadi, maka B kemungkinan besar juga akan terjadi.**"

Analogi yang paling populer untuk menjelaskan ini adalah **Analisis Keranjang Belanja** (*Market Basket Analysis*).





DATA MINING -> ASOSIASI

Penerapan di Dunia Nyata

Association Rule Mining adalah seperti seorang detektif yang mencari petunjuk tersembunyi di dalam data transaksi untuk mengungkap pola perilaku pelanggan.

- **Tata Letak Produk di Toko:** Meletakkan produk yang sering dibeli bersama secara berdekatan (misalnya, keripik di dekat rak minuman soda).
- **Promosi dan Bundling:** Membuat paket promosi seperti "Beli 2 gratis 1" atau "Paket Sarapan" (roti, susu, selai) berdasarkan aturan yang kuat.
- **Rekomendasi Produk:** Mesin rekomendasi di situs e-commerce seperti Tokopedia atau Amazon ("Pelanggan yang membeli barang ini juga membeli...") menggunakan prinsip ini.
- **Deteksi Penipuan:** Menemukan pola transaksi yang tidak biasa dalam data kartu kredit.



Analisis Keranjang Belanja (*Market Basket Analysis*)

Analisis Keranjang Belanja adalah salah satu teknik paling populer dalam data mining yang bertujuan untuk *menemukan hubungan atau asosiasi* antar item-item dalam sebuah dataset transaksi.

Bayangkan kita adalah pemilik supermarket yang ingin tahu: "Jika pelanggan membeli Roti, barang apa lagi yang kemungkinan besar akan mereka beli?"

- **Tata Letak Toko:** Meletakkan produk yang sering dibeli bersama secara berdekatan.
- **Strategi Promosi:** Membuat paket bundling produk (misal: "Beli Roti dan Susu, dapat diskon Mentega").
- **Rekomendasi Produk:** Menampilkan rekomendasi "Pelanggan lain juga membeli..." di toko online.



Konsep Inti: *Itemset* & *Frequent Pattern*

Untuk menemukan hubungan (asosiasi) ini, kita perlu memahami beberapa istilah dasar.

Itemset

- **Definisi:** Sebuah koleksi yang terdiri dari satu atau lebih item. Contoh: {Roti}, {Susu, Mentega}, {Roti, Susu, Telur} adalah beberapa contoh itemset.
- ***k-itemset*:** Sebuah itemset yang berisi k jumlah item. {Susu, Mentega} adalah 2-itemset.

Frequent Pattern / Frequent Itemset

- **Definisi:** Sebuah itemset yang sering muncul dalam dataset. "*Sering*" di sini diukur dengan ambang batas minimum yang kita tentukan, disebut minimum support.
- **Analogi:** Jika kita menetapkan bahwa itemset yang "populer" adalah yang dibeli oleh minimal 30% pelanggan, maka setiap itemset yang memenuhi syarat tersebut adalah *frequent itemset*.

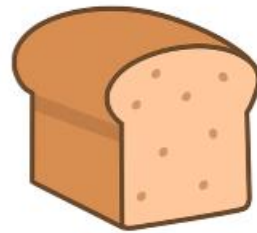


Metrik Pengukuran Asosiasi

Sebuah aturan asosiasi memiliki format Jika **{A}** maka **{B}**, atau dalam notasi: **{A} -> {B}**.

- **Antecedent (A):** Item/itemset di sebelah kiri (Penyebab). Contoh: {Roti}.
- **Consequent (B):** Item/itemset di sebelah kanan (Akibat). Contoh: {Susu}.

Jadi, aturan **{Roti} -> {Susu}** dibaca "*Jika pelanggan membeli Roti, maka mereka juga akan membeli Susu.*"



bread



milk



Metrik Pengukuran Asosiasi

Untuk mengukur seberapa kuat dan menarik sebuah aturan, kita menggunakan **tiga metrik utama**:

1. *Support* (Dukungan)

- **Konsep:** Seberapa populer sebuah itemset. Ini adalah persentase transaksi yang memuat itemset tersebut dari total transaksi.
- **Rumus:**
$$Support(A) = \frac{\text{Jumlah transaksi yang memuat A}}{\text{Total transaksi}}$$
- **Contoh:** Dari 100 transaksi, 30 di antaranya memuat {Roti, Susu}. Maka, $Support(Roti, Susu) = 30/100 = 0.3$ atau 30%.
- **Kegunaan:** Untuk menyaring itemset yang tidak populer. Jika support sebuah itemset sangat rendah, kita tidak perlu repot-repot menganalisisnya lebih lanjut.

2. *Confidence* (Kepercayaan Diri)

- **Konsep:** Seberapa sering aturan $\{A\} \rightarrow \{B\}$ terbukti benar. Ini adalah probabilitas membeli $\{B\}$ jika $\{A\}$ sudah dibeli.
- **Rumus:**
$$Confidence(A \rightarrow B) = \frac{Support(A \cup B)}{Support(A)} = \frac{\text{Jumlah transaksi memuat A dan B}}{\text{Jumlah transaksi memuat A}}$$
- **Contoh:** Jika $Support(Roti, Susu) = 0.3$ dan $Support(Roti) = 0.5$, maka $Confidence(Roti \rightarrow Susu) = 0.3/0.5 = 0.6$ atau 60%. Artinya, 60% dari pelanggan yang membeli Roti, juga membeli Susu.
- **Kelemahan:** Nilai *confidence* bisa tinggi secara menyesatkan jika item B memang sudah sangat populer secara umum.



Metrik Pengukuran Asosiasi

Untuk mengukur seberapa kuat dan menarik sebuah aturan, kita menggunakan **tiga metrik utama**:

3. *Lift* (Daya Angkat)

- **Konsep:** Metrik paling penting untuk mengidentifikasi aturan yang kuat dan menarik. *Lift* mengukur seberapa besar kemungkinan membeli {B} jika {A} dibeli, sambil mempertimbangkan popularitas {B} secara umum.

- **Rumus:**

$$Lift(A \rightarrow B) = \frac{Support(A \cup B)}{Support(A) \times Support(B)} = \frac{Confidence(A \rightarrow B)}{Support(B)}$$

- **Interpretasi:**

- ***Lift* > 1:** Menunjukkan hubungan **positif**. Pelanggan **lebih cenderung** membeli B jika mereka sudah membeli A. Ini adalah aturan yang kita cari! Semakin tinggi nilainya, semakin kuat hubungannya.
- ***Lift* = 1:** Tidak ada hubungan. Pembelian A dan B bersifat independen.
- ***Lift* < 1:** Menunjukkan hubungan **negatif**. Pelanggan **cenderung tidak** membeli B jika mereka sudah membeli A (misalnya, produk substitusi seperti kopi dan teh, atau nasi dan mie instan).



DATA MINING -> ASOSIASI

Algoritma FP-Growth

Untuk menemukan frequent itemsets dari jutaan transaksi, kita memerlukan algoritma yang efisien. Salah satu yang paling populer adalah **FP-Growth**.

- **FP-Growth** adalah singkatan dari Frequent Pattern Growth.
- **Keunggulan** : Jauh lebih cepat dan efisien daripada algoritma pendahulunya (*Apriori*), terutama pada dataset yang besar.

Cara Kerja (FP-Growth)

1. **Scan Dataset Pertama**: Menghitung support untuk setiap item individual dan membuang item yang tidak memenuhi minimum support.
2. **Membangun FP-Tree**: Membuat struktur data berbentuk pohon (tree) yang ringkas untuk merepresentasikan seluruh transaksi. Item yang paling sering muncul akan berada lebih dekat ke akar pohon.
3. **Menambang (Mining) FP-Tree**: Algoritma secara rekursif "menambang" pohon ini untuk menemukan frequent itemsets tanpa harus menghasilkan kandidat itemset berulang kali (seperti yang dilakukan Apriori)



Diskusi Pertemuan 8

Post test pertemuan ke 8 Data mining

Silahkan deskripsikan pemahaman anda terkait metode/algoritma datamining :

Case : Pencarian Data yang berkaitan dengan **Asosiasi** di **Politeknik Negeri Subang**
berikan contoh dan jelaskan alasannya

- Berikan jawaban diskusi di web “<https://learn.sistempolsub.id/moodle/>” pada bagian Pertemuan 11 -> Form Diskusi 11 -> dengan membuat *add discussion* baru. (*cantumkan sumber daftar Pustaka atau referensinya pada isian diskusi*)
- Jika ada hal-hal yang perlu ditanyakan silahkan pada group WA matakuliah