

LECTURE NOTES

COMP8014

Knowledge Data Discovery

Associations Analysis

LEARNING OUTCOMES

LO4: Mahasiswa diharapkan mampu untuk menentukan metode data mining yang sesuai dengan sifat kealamiahannya permasalahan yang dihadapi.

LO5: Mahasiswa diharapkan mampu untuk mengimplementasikan metode data mining.

OUTLINE MATERI :

1. Konsep dasar dan roadmap
2. Metode frequent itemset mining yang efisien dan terukur

ISI MATERI

Apakah Analisis Frequent Pattern?

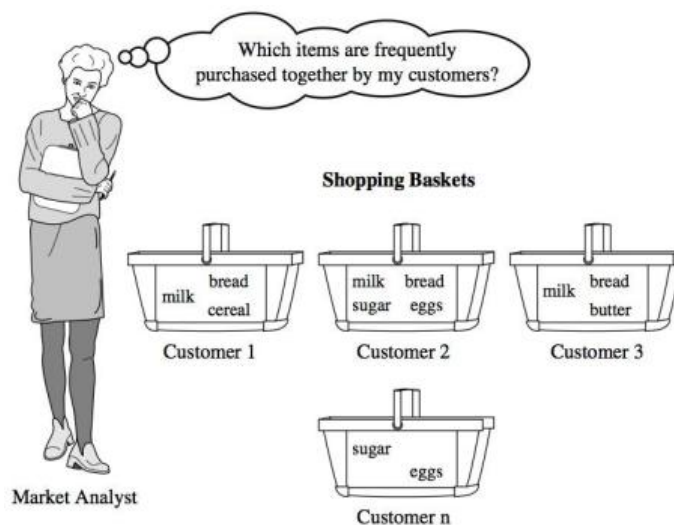
- pola (seperangkat item, subsequences, substructures, dll.) Yang sering terjadi dalam kumpulan data
- Pertama diusulkan oleh Agrawal, Imielinski, dan Swami [AIS93] dalam konteks frequent itemset dan association rule mining

Motivasi dari Frequent Pattern?

- Menemukan keteraturan keteraturan dalam data
 - Produk apa yang sering dibeli bersama? - Bir dan popok ?!
 - Apa pembelian selanjutnya setelah membeli PC?
 - Jenis DNA apa yang sensitif terhadap obat baru ini?
 - Bisakah kita mengklasifikasi dokumen web secara otomatis

Aplikasi dari Frequent Pattern?

Analisa data keranjang, cross-marketing, desain katalog, analisis kampanye penjualan, analisis log web (klik stream), dan analisis urutan DNA.



Transaction-id	Items bought
10	A, B, D
20	A, C, D
30	A, D, E
40	B, E, F
50	B, C, D, E, F

Let $sup_min = 50\%$, $con_fmin = 50\%$

Freq. Pat.: {A:3, B:3, D:4, E:3, AD:3}

Association rules:

$A \Rightarrow D$ (sup=60%, conf=100%)

$D \Rightarrow A$ (sup=60%, conf=75%)

$P(A \wedge D)$ = jumlah transaksi yang mengandung A dan D / jumlah transaksi = $3/5 = 60\%$

$P(D|A)$ = jumlah transaksi yang mengandung A dan D / jumlah transaksi A = $3/3 = 100\%$

$P(D \wedge A)$ = jumlah transaksi yang mengandung A dan D / jumlah transaksi = $3/5 = 60\%$

$P(A|D)$ = jumlah transaksi yang mengandung A dan D / jumlah transaksi D = $3/4 = 75\%$

SIMPULAN

1. Frequent pattern adalah pola (satu set item, subsequences, substructures, dll.) Yang sering terjadi pada kumpulan data.
2. Frequent pattern mengungkapkan properti intrinsik dan penting dari kumpulan data.
3. Suatu downward closure property dari frequent pattern melibatkan setiap subset dari frequent itemset harus sering dilakukan
4. Tiga pendekatan utama dalam metode penambangan Scalable: Apriori (Agrawal & Srikant @ VLDB'94), Frek. Pola pertumbuhan (FPgrowth-Han, Pei & Yin @ SIGMOD'00) dan pendekatan format data vertikal (Charm-Zaki & Hsiao @ SDM'02).
5. Aturan asosiasi bisa dihasilkan dari Frequent Itemsets.

DAFTAR PUSTAKA

1. Han, J., Kamber, M., & Pei, Y. (2006). “Data Mining: Concepts and Technique”. Edisi 3. Morgan Kaufman. San Francisco
2. Tan, P.N., Steinbach, M., & Kumar, V. (2006). “Introduction to Data Mining”. Addison-Wesley. Michigan
3. Witten, I. H., & Frank, E. (2005). “Data Mining : Practical Machine Learning Tools and Techniques”. Second edition. Morgan Kaufmann. San Francisco