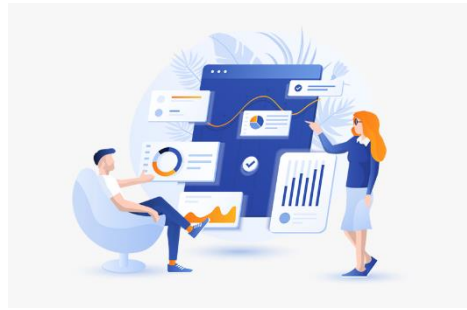


DATA MINING

13 – Algoritma Apriori

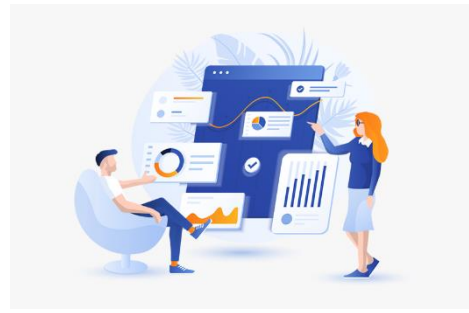
Oleh: Leny Tritanto N., M.Kom.

Sumber: (Kenali Data Science
Algoritma Apriori pada Machine
Learning
, <https://dqlab.id/>)



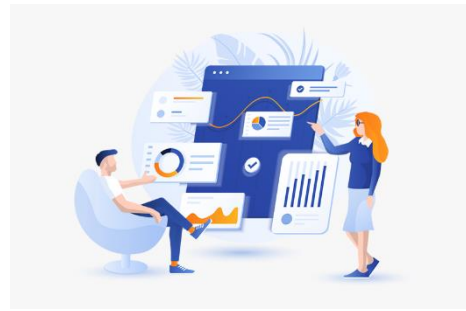
Algoritma Supervised dan Unsupervised Learning:

Supervised Learning	Unsupervised Learning
▪ Linear Regression	✓ K-Means
✓ Decision Tree and Random Forest	✓ Hierarchical Clustering
✓ Naive Bayes Classifier	✓ DBSCAN
✓ Nearest Neighbour Classifier (KNN)	✓ Association Rule
✓ Artificial Neural Network	➡ Apriori Algorithm
✓ Support Vector Machine (SVM)	



Apa dan bagaimana
menerapkan Algoritma
Apriori?





Contoh Implementasi

Minimarket / Supermarket



Toko Buku / Perpustakaan

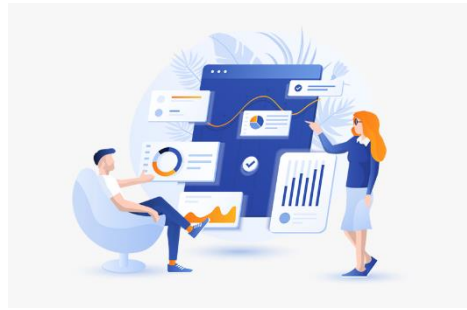


Toko Handphone



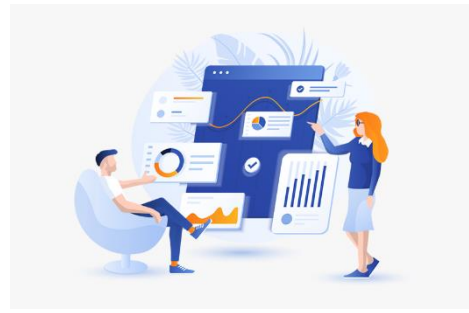
Paket Wisata





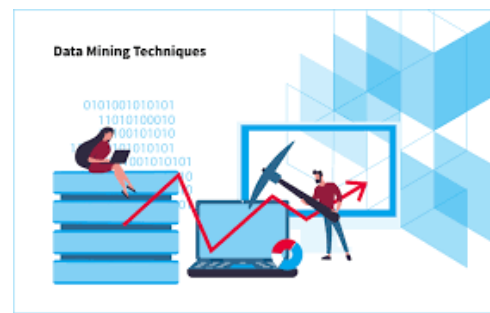
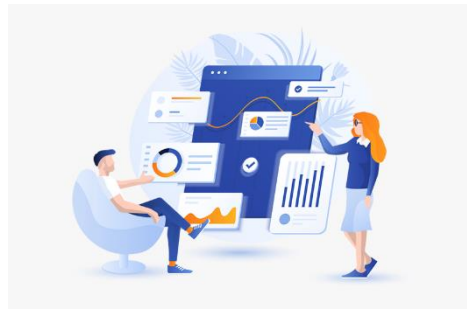
Algoritma Apriori

- Algoritma apriori adalah salah satu algoritma pada data mining untuk mencari *frequent* (sering) item atau itemset pada transaksional database.
- Algoritma apriori banyak digunakan pada data transaksi atau biasa disebut dengan **market basket**, misalnya sebuah swalayan atau minimarket memiliki market basket (data transaksi), dengan algoritma apriori minimarket tersebut dapat mengetahui pola pembelian *customer*.
- Algoritma apriori juga dapat digunakan oleh restoran cepat saji dalam menentukan paket menu makanan atau minuman yang ada di restoran berdasarkan pola beli konsumen.



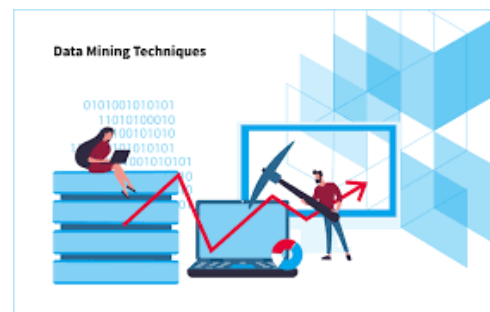
Manfaat Algoritma Apriori

1. Untuk *market basket analysis* (menganalisa kebiasaan customer dalam memilih item pembelian)
2. Sistem Rekomendasi
3. Menyediakan stok
4. Pemberian diskon
5. Sistem bundling



Konsep Algoritma Apriori

Metode data mining digunakan untuk mencari frekuensi hubungan atau pola hubungan antar item dalam dataset



Indikator dalam Algoritma Apriori



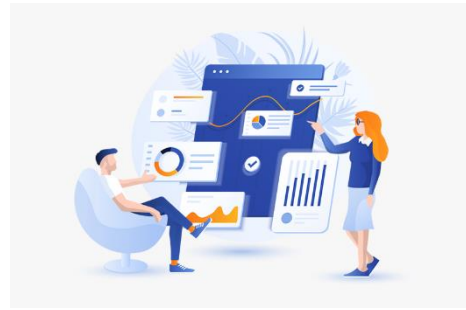
SUPPORT / Penunjang : Persentase kombinasi item dalam suatu dataset transaksi



CONFIDENCE / Kepastian : Persentase kuatnya hubungan antar item dalam aturan asosiatif



LIFT RATIO : Persentase hasil uji validasi rule yang dihasilkan



Hal penting antara SUPPORT, CONFIDENCE, dan LIFT RATIO



Menghitung **SUPPORT** ($A \cup B$) : Jumlah Transaksi mengandung A dan B
Total Transaksi



CONFIDENCE ($A \rightarrow B$) : Support Count ($A \cup B$)
Support Count (A)



LIFT RATIO : Confidence
Benchmark Confidence

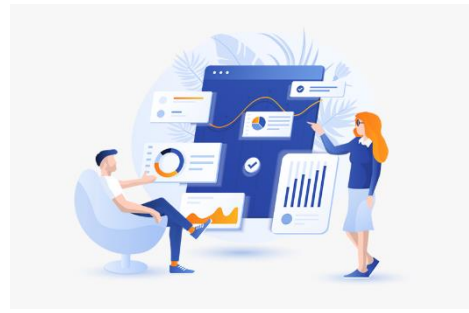


BENCHMARK CONFIDENCE : Jumlah transaksi dengan item *consequent*
Jumlah Transaksi

Note:

Jika ada himpunan transaksi A dan B maka item yang mendahului / **antecedent** adalah **A** dan item yang mengikuti / **consequent** adalah **B**

A (antecedent) \rightarrow B(Consequent)



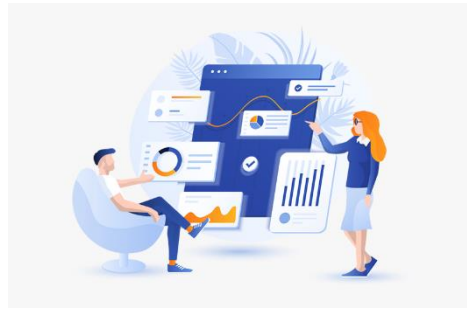
Istilah dalam Algoritma Asosiasi / Apriori

1. E adalah himpunan. Contoh: {asparagus, beans, potatoes,...,tomatoes}
2. D adalah jumlah seluruh transaksi. Contoh: {transaksi1, transaksi2,..., transaksi-n}
3. Proper subset adalah himpunan bagian murni. Contoh:
 - a. Ada suatu himpunan $A = \{a, b, c\}$
 - b. Himpunan bagian dari A adalah
 - Himpunan kosong = $\{\}$
 - Himpunan 1 unsur = $\{a\}, \{b\}, \{c\}$
 - Himpunan 2 unsur = $\{a, b\}, \{b, c\}, \{a, c\}$
 - Himpunan 3 unsur = $\{a, b, c\}$



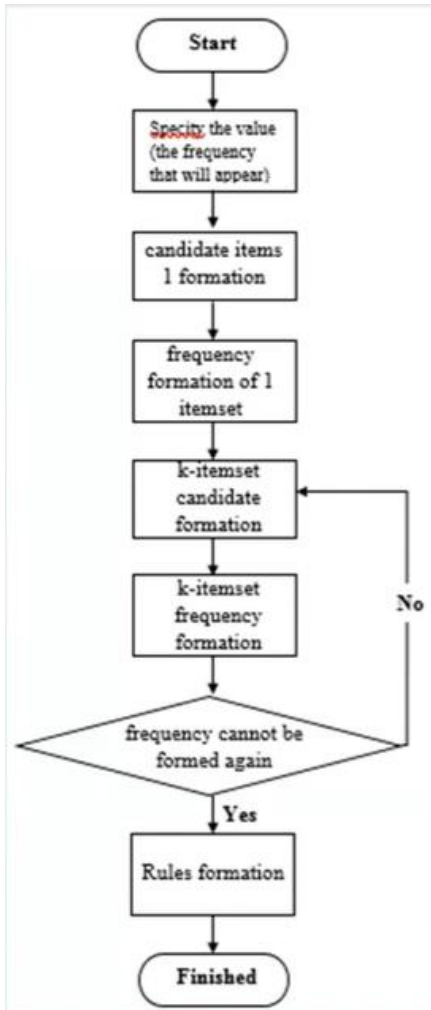
Istilah dalam Algoritma Asosiasi / Apriori

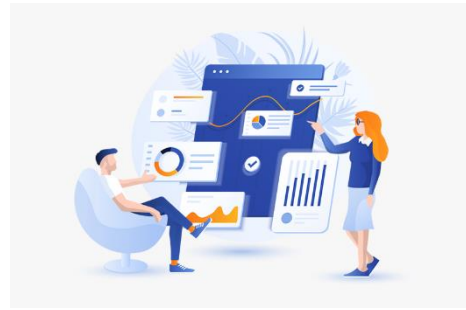
4. Itemset adalah item-item di E.
5. K-Itemset adalah itemset yang terdiri dari K buah item yang ada pada E. K adalah jumlah unsur yang terdapat pada suatu himpunan. Contoh: 2-itemset adalah yang bersifat 2 unsur {a,b}, {b,c}, {a,c}
6. Itemset frekuensi adalah jumlah transaksi yang mengandung jumlah item set tertentu (jumlah transaksi terhadap itemset yang dibeli). Contoh:
 - Frekuensi item set yang membeli kopi dan gula adalah 3
 - Frekuensi item set yang membeli kopi, gula, dan susu adalah 2
7. Frekuensi itemset adalah itemset yang muncul sekurang-kurangnya “sekian” kali D(minimum support). Kata sekian disimbolkan dengan θ (Theta). θ merupakan batas minimum / **threshold** dalam suatu transaksi.



Flowchart Algoritma Apriori

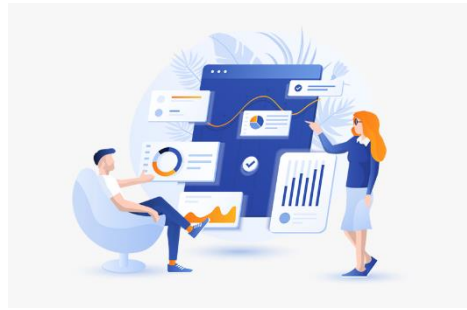
1. Tentukan nilai minimum support dan confidence
2. Spesifikasikan semua itemset dalam transaksi dan hitung frekuensi kemunculannya
3. Seleksi itemset berdasarkan threshold / min support item (poin 1). Item yang jumlah frekuensinya tidak sesuai min support akan dihilangkan.
4. Bentuk K Itemset (kombinasi 2 itemset terlebih dahulu)
5. Hitung kembali frekuensi k-itemset dan dipilih minimal threshold/ min support
6. Pada poin 5 akan terbentuk 2 itemset, langkah selanjutnya bentuk 3 itemset.
7. Pembentukan k-itemset bisa dilanjutkan sampai seterusnya. Pembentukan k-itemset berhenti ketika sudah tidak memenuhi ambang threshold
8. Rule terbentuk dan hitung nilai confidence.





Cara Kerja Algoritma Apriori

1. Tentukan minimum support dan confidence untuk mengetahui penting atau tidaknya suatu aturan asosiatif
2. Iterasi 1: hitung item-item dari support (transaksi yang memuat seluruh item) dengan menelusuri database untuk 1 itemset, setelah 1-itemset didapatkan, dari 1-itemset apakah diatas minimum support, apabila telah memenuhi minimum support, i-itemset tersebut akan menjadi pola frequent tinggi
3. Iterasi 2: untuk mendapatkan 2-itemset, harus dilakukan kombinasi dari k-itemset sebelumnya, kemudian telusuri database kembali untuk menghitung item-item yang memuat support. Itemset yang memenuhi minimal support akan dipilih sebagai pola frequent tinggi
4. Tetapkan nilai k-itemset dari support yang telah memenuhi minimum support dari k-itemset
5. Lakukan proses untuk iterasi selanjutnya hingga tidak ada lagi k-itemset yang memenuhi minimum support



Cara Kerja Algoritma Apriori

Misal, terdapat data transaksi sebagai berikut:

$(\text{roti, mentega}) \rightarrow \{\text{susu}\}$ (support = 40% dan confidence = 50%)

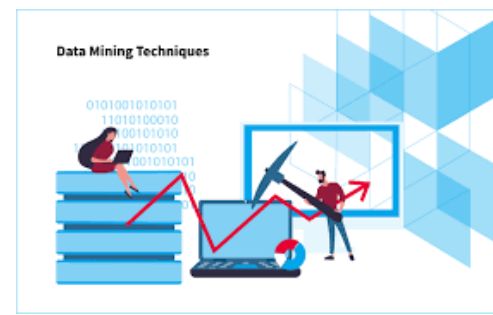
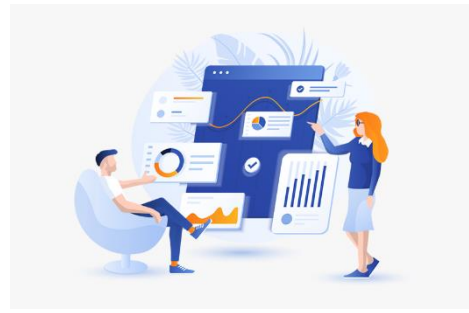
Artinya:

- 40% dari seluruh transaksi dalam database memuat ketiga item tersebut
- 50% dari transaksi database memuat item roti dan mentega juga membeli susu
- “Seorang pelanggan yang membeli roti dan mentega memiliki kemungkinan 50% untuk membeli susu juga.

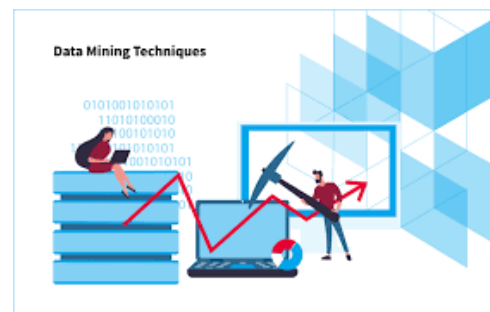
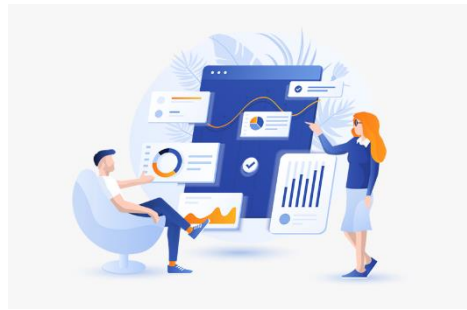
Note:

Support (nilai penunjang) = presentase kombinasi item dalam database

Confidence (nilai kepastian) = kuatnya hubungan antar item dalam aturan asosiasi



CONTOH KASUS



Contoh Kasus

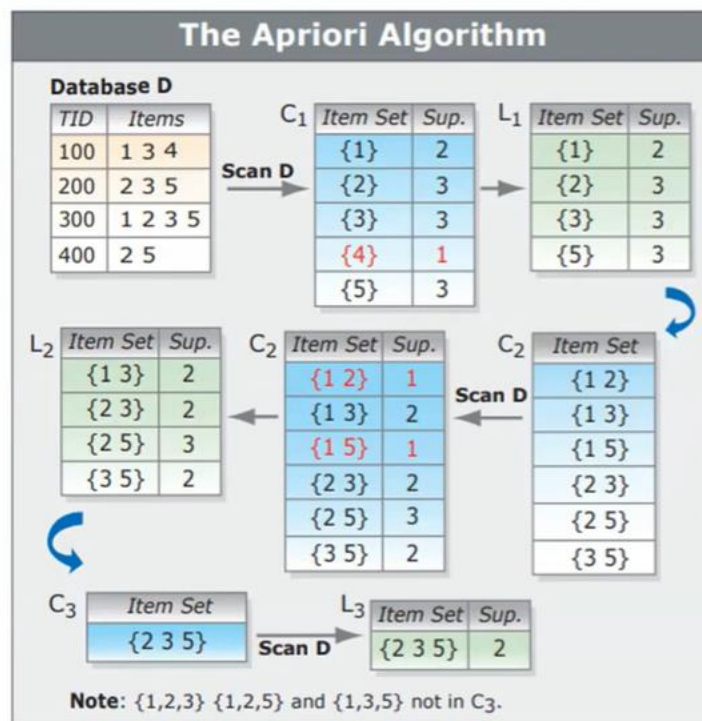


Image by MIT OpenCourseWare, adapted from Osmar R. Zaiane.

PROSES APRIORI

contoh: MIN SUPPORT = 2


Confident = 60%

PENTING:

Jika dalam sebuah transaksi ke-n, item x dibeli 2x atau lebih dalam 1 transaksi, maka tetap dihitung 1




Contoh Kasus



Database D
Database D

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5



Item Set	Sup.
{2 3 5}	2

Misal: Min Confidence 60%

Nilai Support = 2 / Threshold/Φ

– Confidence

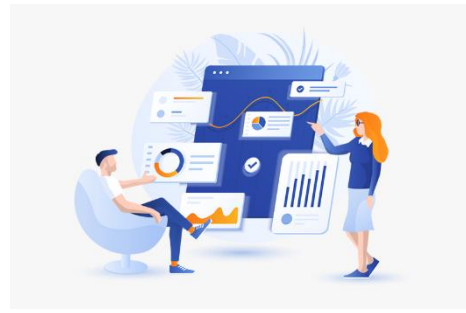
$$\hat{P}(b|a) = \frac{\text{Supp}(a \cup b)}{\text{Supp}(a)}$$

A(antecedent) → B(Consequence)


Confidence (A → B) = Support_count (A ∪ B) / Support_count (A)

PEMBUATAN RULE L-3/ 3 Itemset kombinasi

- Conf {2,3} → {5} = 2/2 * 100% = 100%
- Conf {2,5} → {3} = 2/3 * 100% = 66,66%
- Conf {3,5} → {2} = 2/2 * 100% = 100%
- Conf {2} → {3,5} = 2/3 * 100% = 66,66%
- Conf {3} → {2,5} = 2/3 * 100% = 66,66%
- Conf {5} → {2,3} = 2/3 * 100% = 66,66%




Contoh Kasus



Database D
Database D

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5



Item Set	Sup.
{2 3 5}	2

Misal: Min Confidence 60%

Nilai Support = 2/ Threshold/ Φ

– Confidence

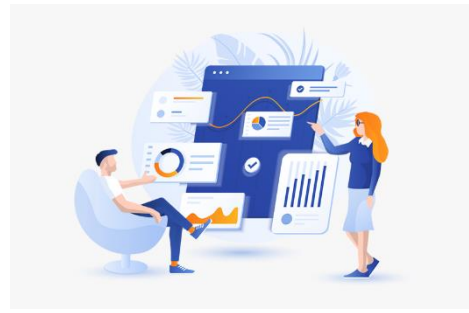
$$\hat{P}(b|a) = \frac{\text{Supp}(a \cup b)}{\text{Supp}(a)}$$

A(antecedent) → B(Consequence)

Confidence (A → B) = $\text{Support_count}(A \cup B) / \text{Support_count}(A)$

PEMBUATAN RULE L-3/ 3 Itemset kombinasi

- Conf {2,3} → {5} = $2/2 \times 100\% = 100\%$
- Conf {2,5} → {3} = $2/3 \times 100\% = 66,66\%$
- Conf {3,5} → {2} = $2/2 \times 100\% = 100\%$
- Conf {2} → {3,5} = $2/3 \times 100\% = 66,66\%$
- Conf {3} → {2,5} = $2/3 \times 100\% = 66,66\%$
- Conf {5} → {2,3} = $2/3 \times 100\% = 66,66\%$



Contoh teknik membaca rule dari hasil 3 itemset

1. Confidence itemset $\{2,3\} \rightarrow \{5\}$, artinya: jika item 2 dan 3 dibeli, maka customer membeli item 5 dengan tingkat kepercayaan / confidence 100%.
2. Contoh 1 \rightarrow strategi penerapan rule pada minimarket : pembelian item 2 dan 3 akan mendapatkan bonus item 5 dengan strategi harga dinaikan.
3. Contoh 2 \rightarrow peletakan barang item 2 dekat dengan item 3 kemudian disusun dekat rak barang item 5.