

# **TUGAS AKHIR DATA MINING**



Nama Lengkap : Dimas Tsaqif Aulia  
NIM : A11.2020.12790

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS DIAN NUSWANTORO  
TAHUN AJARAN  
2023**

## EKSPERIMEN K-MEDOIDS UNTUK CLUSTERING DATA DAN FP-GROWTH UNTUK PROSES ASOSIASI

Dalam dataset ini mendefinisikan transaksi penjualan pada sebuah supermarket. Terdapat 304 atribut yang 303 merupakan item/barang dan salah satunya merupakan ID transaksi penjualan. Terdiri dari 1361 record transaksi. Berikut adalah contoh dataset sebagai berikut:

Basket id	Lemon	Standardcoffee	FrozenChickenWings	98pct. Fat Free Hamburger	Sugar Cookies	Onions
C11867	false	false	true	true	false	false
C5096	false	false	true	true	false	true
C4295	false	false	false	false	false	true
C2837	true	false	false	false	true	false
C2693	true	false	true	false	true	true
C3497	false	true	false	true	false	false

### Keterangan:

1. **Basket id** merupakan ID Transaksi penjualan barang.
2. **Lemons, Standard coffee, Frozen Chicken Wings, 98pct. Fat Free Hamburger, Sugar Cookies, dan Onions** merupakan barang-barang yang dijual di Supermarket.
3. **False** menandakan bahwa barang tersebut tidak dibeli oleh pembeli.
4. **True** menandakan bahwa barang tersebut dibeli oleh pembeli.

### 1. Pengolahan Data Awal

Data yang sudah dikumpulkan akan diolah dengan beberapa tahap sehingga menjadi sebuah dataset yang akan diimplementasikan terhadap metode-metode data mining. Semua record transaksi yang ada pada dataset penjualan digunakan dalam tahap pengolahan data. Tahapan pengolahan data yang dilakukan yaitu:

#### a. Data Preprocessing

Data Preprocessing yaitu dilakukan pembersihan dan persiapan data untuk menghilangkan konsistensi data, data tidak lengkap dan redundant data yang terdapat pada data awal. Data Preprocessing juga melakukan perubahan status barang yang semula true / false menjadi 1 / 0, yang akan digunakan untuk proses asosiasi.

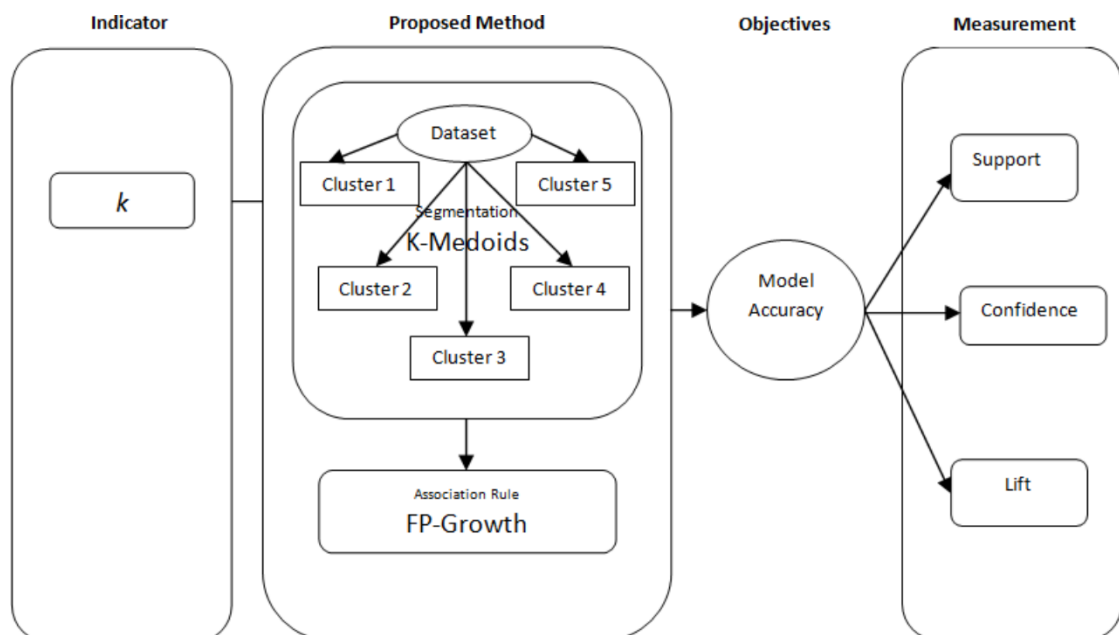
#### b. Menghitung Jumlah Barang yang Dibeli / Tidak Dibeli

Penghitungan jumlah barang yang dibeli maupun yang tidak dibeli dilakukan dengan menambah 2 atribut di dalam dataset, kedua atribut tersebut digunakan untuk melakukan proses clustering data.

Basket ID	Jumlah Barang yang dibeli	Jumlah Barang yang tak dibeli	Hair Conditioner	Lemons	Standard coffee
C11867	1	302	0	0	0
C5096	2	301	0	0	0
C4295	1	302	0	0	0
C2837	1	302	0	0	0
C2693	1	302	0	0	0
C3497	1	302	0	0	0
C2696	1	302	0	0	0
C1895	17	286	0	0	0
C9524	1	302	0	0	0

## 2. Desain Eksperimen

Eksperimen pada penelitian ini adalah penggunaan *K-Medoids* untuk *clustering* data dan *FP-Growth* untuk proses asosiasi.



Eksperimen menggunakan algoritma *K-Medoids* untuk *clustering* pada dataset dan menerapkan algoritma *FP-Growth* untuk pendekatan asosiasi pada setiap *cluster*. Sebelum data yang sudah dikumpulkan diolah untuk menemukan pola asosiasi dengan menggunakan *FP-Growth*, terlebih dahulu dilakukan proses *clustering* data dengan menggunakan *K-Medoids*. *Clustering* data dilakukan dengan cara membagi data yang ada menjadi 5 *cluster* (k) berdasarkan jumlah barang yang dibeli maupun yang tidak dibeli. Tujuan dari *clustering* adalah agar data yang di asosiasi menjadi lebih kecil sehingga pola yang dihasilkan dapat lebih akurat.

Masing-masing dari *cluster* yang terbentuk akan dilakukan proses asosiasi menggunakan algoritma *FP-Growth* untuk menentukan rekomendasi produk kepada pelanggan. Hasil dari proses asosiasi ini diukur menggunakan *Support*, *Confidence*, *Lift Ratio*. Akurasi dari penerapan algoritma *FP-Growth* yang di *clustering* dahulu menggunakan *K-Medoids* akan dibandingkan dengan akurasi dari penerapan algoritma *FP-Growth* saja.

### 3. Evaluasi Penelitian

Hasil akurasi untuk menentukan rekomendasi produk kepada pelanggan yang didapat dari penelitian ini dapat diukur dengan menggunakan *Lift Ratio*. *Lift Ratio* merupakan nilai yang menunjukkan kevalidan proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. Sebuah transaksi dikatakan valid jika mempunyai nilai *Lift Ratio* lebih dari 1, yang berarti bahwa dalam transaksi tersebut, produk A dan B benar-benar dibeli secara bersamaan. *Lift Ratio* mengukur seberapa penting *rule* yang telah terbentuk berdasarkan nilai *support* dan *confidence*.

No.	Premises	Conclusion	Support	Confidence	Lift
1	Whole Corn	Hair Conditioner	0.970	0.976	1.006
2	Orange Flavored Fruit Bars	Hair Conditioner	0.970	0.976	1.006
3	Whole Corn	Orange Flavored Fruit Bars, Hair Conditioner	0.970	0.976	1.006
4	Orange Flavored Fruit Bars	Whole Corn, Hair Conditioner	0.970	0.976	1.006
5	Whole Corn, Orange Flavored Fruit Bars	Hair Conditioner	0.970	0.976	1.006
6	Whole Corn	Orange Flavored Fruit Bars	0.994	1	1.006
7	Orange Flavored Fruit Bars	Whole Corn	0.994	1	1.006
8	Hair Conditioner	Whole Corn	0.970	1	1.006
9	Hair Conditioner	Orange Flavored Fruit Bars	0.970	1	1.006
10	Hair Conditioner	Whole Corn, Orange Flavored Fruit Bars	0.970	1	1.006
11	Whole Corn, Hair Conditioner	Orange Flavored Fruit Bars	0.970	1	1.006
12	Orange Flavored Fruit Bars, Hair Conditioner	Whole Corn	0.970	1	1.006

