

LKP 9 DATA MINING

Nama : Shibgotalloh Sabilana

NIM : G64180002

1. Buka data data9.csv (ipb.link/datasetlkp9)
2. Gunakan algoritme Apriori dan FPGrowth untuk menemukan aturan asosiasi masing masing algoritme dari dataset, jelaskan hasil yang diperoleh!

• Algoritma Apriori

```
Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     Data9
Instances:    100
Attributes:   7
              Home
              Browsed
              Searched
              Prod_A
              Prod_B
              Prod_C
              Visit_Again
=== Associator model (full training set) ===
```

Pada data9.csv ini terdiri dari 100 contoh sampel data yang terdiri dari 7 atribut (Home, Browsed, Searched, Prod_A, Prod_B, Prod_C, Visit_again)

```
Minimum support: 0.2 (20 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16
```

Generated sets of large itemsets:

Size of set of large itemsets L(1): 14

Size of set of large itemsets L(2): 61

Size of set of large itemsets L(3): 44

Size of set of large itemsets L(4): 7

Pada proses Apriori ini menetapkan minimum supportnya sebesar 0.2 dan minimum metric sebesar 0.9. Lalu dapat terlihat hasil dari proses tersebut diperoleh :

- ❖ Saat kita men-generated 1 item set, terdapat 14 aturan asosiasi yang didapatkan
- ❖ Saat men-generated 2 item set, terdapat 61 aturan asosiasi yang didapatkan
- ❖ Saat men-generated 3 item set, terdapat 44 aturan asosiasi yang didapatkan
- ❖ Saat men-generated 4 item set, terdapat 7 aturan asosiasi yang didapatkan

Best rules found:

```
1. Home=no Prod_A=no 28 ==> Browsed=yes 28 <conf:(1)> lift:(1.39) lev:(0.08) [7] conv:(7.84)
2. Home=no Prod_A=no 28 ==> Prod_B=yes 28 <conf:(1)> lift:(1.82) lev:(0.13) [12] conv:(12.6)
3. Home=no Prod_A=no Prod_B=yes 28 ==> Browsed=yes 28 <conf:(1)> lift:(1.39) lev:(0.08) [7] conv:(7.84)
4. Home=no Browsed=yes Prod_A=no 28 ==> Prod_B=yes 28 <conf:(1)> lift:(1.82) lev:(0.13) [12] conv:(12.6)
5. Home=no Prod_A=no 28 ==> Browsed=yes Prod_B=yes 28 <conf:(1)> lift:(2.17) lev:(0.15) [15] conv:(15.12)
6. Home=no Prod_C=yes 21 ==> Browsed=yes 21 <conf:(1)> lift:(1.39) lev:(0.06) [5] conv:(5.88)
7. Prod_A=no Prod_B=yes Prod_C=yes 21 ==> Browsed=yes 21 <conf:(1)> lift:(1.39) lev:(0.06) [5] conv:(5.88)
8. Searched=no Prod_A=no Prod_B=yes 20 ==> Browsed=yes 20 <conf:(1)> lift:(1.39) lev:(0.06) [5] conv:(5.6)
9. Prod_A=no Prod_B=yes 34 ==> Browsed=yes 33 <conf:(0.97)> lift:(1.35) lev:(0.09) [8] conv:(4.76)
10. Home=no Browsed=yes Prod_B=yes 29 ==> Prod_A=no 28 <conf:(0.97)> lift:(2.05) lev:(0.14) [14] conv:(7.68)
```

Disini dihasilkan rules yang dimunculkan sebanyak 10, karena saat sebelum proses telah ditetapkan jika NumRules = 10

- Algoritma FPGrowth

```
Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation:     Data9
Instances:    100
Attributes:   7
```

```
Home
Browsed
Searched
Prod_A
Prod_B
Prod_C
Visit_Again
```

```
=== Associator model (full training set) ===
```

```
FPGrowth found 7 rules (displaying top 7)
```

```
1. [Prod_C=yes, Home=no]: 21 ==> [Browsed=yes]: 21 <conf:(1)> lift:(1.39) lev:(0.06) conv:(5.88)
2. [Prod_B=yes, Prod_C=yes, Home=no]: 19 ==> [Browsed=yes]: 19 <conf:(1)> lift:(1.39) lev:(0.05) conv:(5.32)
3. [Searched=yes, Home=no]: 12 ==> [Browsed=yes]: 11 <conf:(0.92)> lift:(1.27) lev:(0.02) conv:(1.68)
4. [Browsed=yes, Home=no]: 32 ==> [Prod_B=yes]: 29 <conf:(0.91)> lift:(1.65) lev:(0.11) conv:(3.6)
5. [Prod_C=yes, Home=no]: 21 ==> [Prod_B=yes]: 19 <conf:(0.9)> lift:(1.65) lev:(0.07) conv:(3.15)
6. [Prod_C=yes, Home=no]: 21 ==> [Browsed=yes, Prod_B=yes]: 19 <conf:(0.9)> lift:(1.97) lev:(0.09) conv:(3.78)
7. [Browsed=yes, Prod_C=yes, Home=no]: 21 ==> [Prod_B=yes]: 19 <conf:(0.9)> lift:(1.65) lev:(0.07) conv:(3.15)
```

Karena pada FPGrowth tidak ada nilai cycle nya maka hasil yang didapatkan langsung memunculkan best rules yang ditemukan. Disini diperoleh terdapat 7 rules yang ditemukan

3. Ukuran kemenarikan apa saja yang digunakan kedua algoritme tersebut? Jelaskan!

Pada proses baik algoritma apriori maupun algoritma FPGrowth terdapat dua parameter yang menjadi ukuran ketertarikan pada proses asosiasi yaitu support dan confidence

4. Apakah pengaruh nilai minimum support dan confidence terhadap aturan yang diperoleh? Apakah maksud dari aturan dengan nilai support dan confidence yang tinggi? Jelaskan!

Minimum support dibuat untuk menjadi batas minimal setiap frequent itemset yang

muncul. Contohnya frequent item set merupakan itemset yang muncul minimal sekali di minimum support. Sehingga semakin besar nilai minimum supportnya otomatis semakin sedikit nilai frequent yang muncul

Confident merupakan kuatnya hubungan antar item pada fungsi asosiasi, nilai ini juga berpengaruh terhadap banyaknya rules yang ditemukan. Oleh karena itu semakin kecil nilai confidence yang ditetapkan maka semakin banyak kemungkinan rules yang diperoleh. Namun tentunya nilai minimum confidence yang terlalu kecil tidak baik juga terhadap proses yang dijalankan

5. Untuk data set seperti apakah algoritme Apriori dan FPGrowth cocok digunakan? Mengapa data seperti diabetes.arff dan iris.arff tidak dapat dilakukan analisis asosiasi, terutama dengan menggunakan Weka?

Yaitu pada dataset yang merupakan data-data transaksi alias market based. Nahh data transaksi tersebut memiliki kategori nilai dalam bentuk kategorik. Lalu kenapa data seperti diabetes.arff dan iris.arff tidak dapat dilakukan analisis asosiasi? Karena pada data-data tersebut kategori nilai yang dipakai adalah numerik bukan kategorik, sehingga dataset tersebut tidak dapat dilakukan proses analisis asosiasi