PRAKTIKUM DATA WAREHOUSING DAN DATA MINING MODUL 12 INDUKSI DAN ATURAN ASOSIASI



Disusun oleh: Adinda Aulia Hapsari L200220037

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA
TAHUN 2024

Setelah kegiatan selesai, lembar kerja ini dicetak (di-print) dan dikumpulkan ke

asisten.

NIM : L200220037

Nama : Adinda Aulia Hapsari

Nama Asisten : Diva Halimah Tanggal Praktikum : 21 Desember 2024 (Diisi oleh Asisten)

Nilai Praktek:

Tanda Tangan:

KEGIATAN PRAKTIKUM

12.4.1 Mengimport Library

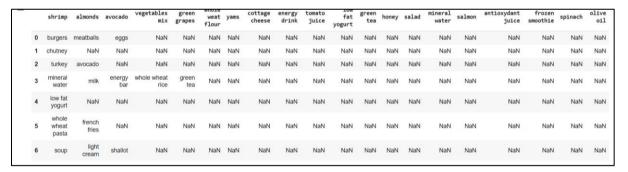
Meng-import library yang diperlukan, yaitu library pandas, numpy, dan apriori. Untuk meng-import library yang akan digunakan, kita meng import library pandas, numpy, dan apriori.

```
[ ] import numpy as np
import pandas as pd
from apyori import apriori
```

12.4.2 Membaca Dataset

Membaca dataset dari sebuah directory. Disini, kita mengambil dataset store_data.csv yang diambil dari platform gitea. Dataset tersebut akan tersimpan di variable store_data sebagai sebuah dataframe. Kemudian, kita akan mencetak tujuh data teratas pada dataset.





12.4.3 Mengkonversikan Dataframe ke dalam Array

Pada kode di bawah ini, kita akan mengkonversikan dataframe menjadi sebuah array bernama records. Disini, kita mengaplikasikan for loop dimulai dari baris pertama hingga terakhir. Kemudian di setiap iterasi tersebut, kita membuat sebuah list yang menyimpan nilai kolom pada setiap barisnya. Terakhir, kita cetak array records tersebut.

```
records = []
for i in range(0, store_data.shape[0]):
    records.append([str(store_data.values[i, j]) for j in range(0, store_data.shape[1])])
print(records)

[['burgers', 'meatballs', 'eggs', 'nan', 'nan',
```

12.4.4 Membuat Model Aturan Asosiasi

Selanjutnya, kita jalankan kode untuk aturan asosiasi menggunakan algoritma apriori dengan cara memanggil fungsi apriori. Adapun fungsi apriori memiliki beberapa parameter sebagai berikut:

- 1. Parameter pertama adalah array yang akan digunakan yaitu records.
- 2. Parameter kedua adalah nilai minimum support (min_support=0.0045) yang berarti nilai minimum confidence yang diimplementasikan sebagai ambang batas confidence adalah 0.45%
- 3. Parameter ketiga adalah minimum confidence (min_confidence=0.2) yang berarti nilai minimum confidence yang diimplementasikan sebagai ambang batas confidence adalah 20%
- 4. Parameter keempat adalah minimum lift ratio (min_lift = 3) yang berarti nilai minimum lift ratio yang diimplementasikan sebagai ambang batas lift ratio adalah 3.
- 5. Parameter kelima adalah min_length = 2 adalah jumlah minimum item yang kita inginkan dalam aturan asosiasi. Disini kita menggunakan min_length = 2 berarti kita menginginkan setidaknya dua produk dalam aturan asosiasi yang dihasilkan.

Hasil dari fungsi apriori disimpan ke dalam variabel association_ rules, kemudian association_results akan menyimpan association_ rules yang telah diubah menjadi sebuah list. Terakhir, kita cetak nilai association results dengan menggunakan iterasi.

telationRecord(items=frozenset(['chicken', 'light cream']), support=0.0045333333333333, ordered statistics=[orderedStatistic(items base=frozenset(['light cream']), items add=frozenset(['chicken', 'light cream']), items add=frozenset(telationRecord(items-frozenset({'mushroom cream sauce', 'escalope'}), support=0.005733333333333, ordered_statistics=[OrderedStatistic(items_base=frozenset({'mushroom cream sauce'}), items_add=f kelationRecord(items=frozenset(['pasta', 'escalope'}), support=0.00588666666666667, ordered_statistics=[OrderedStatistic(items_base=frozenset(['pasta']), items_add=frozenset(['escalope']), confide telationRecord(items=frozenset({'ground beef', 'herb & pepper'}), support=0.016, ordered statistics=[OrderedStatistic(items base=frozenset(('herb & pepper')), items add=frozenset(('ground beef')), RelationRecord(items=frozenset({'ground beef', 'tomato sauce'}), support=0.005333333333333, ordered statistics=[OrderedStatistic(items base=frozenset({'tomato sauce'}), items add=frozenset({'gr telationRecord(items=frozenset({'whole wheat pasta', 'olive oil'}), support=0.008. ordered statistics=[OrderedStatistic(items base=frozenset({'whole wheat pasta'}), items add=frozenset({'olive oil telationRecord(items=frozenset(('shrimp', 'pasta')), support=0.00506666666666666666, ordered statistics=[Orderedstatistic(items base=frozenset(('pasta')), items add=frozenset(('shrimp'), confidence RelationRecord(items=frozenset({'nan', 'chicken', 'light cream'}), support=0.0045333333333333334, ordered_statistics=[OrderedStatistic(items_base=frozenset({'ir cord(items=frozenset({'shrimp', 'frozen vegetables', 'chocolate'}), support=0.005333333333333, ordered_statistics=[OrderedStatistic(items_base=frozenset({'frozen vegetables', 'chocolate'}) telationRecord(items=frozenset({'ground beef', 'cooking oil', 'spaghetti'}}), support=0.0048, ordered_statistics=[OrderedStatistic(items_base=frozenset({'ground beef', 'cooking oil'}), items_add=fr telationRecord(items-frozenset({'nan', 'mushroom cream sauce', 'escalope'}), support=0.0057333333333333, ordered statistics=[OrderedStatistic(items base=frozenset({'mushroom cream sauce'}), item RelationRecord(items=frozenset(('nan', 'pasta', 'escalope')), support=0.0058666666666667, ordered_statistics=[OrderedStatistic(items_base=frozenset(('pasta')), items_add=frozenset(('nan', 'escalo RelationRecord(items=frozenset({'ground beef', 'frozen vegetables', 'spaghetti'}), support=0.008666666666666666666666. ordered statistics=[OrderedStatistic(items base=frozenset({'frozen vegetables', 'spaghetti'})] nRecord(items=frozenset({'milk', 'frozen vegetables', 'olive oil'}), support=0.0048, ordered statistics=[orderedStatistic(items base=frozenset({'milk', 'frozen vegetables'}), items add=fro: RelationRecord(items=frozenset(('minera) water', 'frozen vegetables', 'shrimn')), support=0.0072, ordered statistics=[OrderedStatistic(items base=frozenset(('minera) water', 'shrimn')), items add=f telationRecord(items=frozenset({'frozen vegetables', 'olive oil', 'spaghetti'}), support=0.005733333333333, ordered statistics=[OrderedStatistic(items base=frozenset({'frozen vegetables', 'spaghetti'})] RelationRecord(items-frozenset({'shrimp', 'frozen vegetables', 'spaghetti'}), support=0.006, ordered statistics=[OrderedStatistic(items base=frozenset({'frozen vegetables', 'spaghetti'}), items add rd(items=frozenset({'frozen vegetables', 'tomatoes', 'spaghetti'}), support=0.006666666666667, ordered_statistics=[OrderedStatistic(items_base=frozenset({'frozen vegetables', 'spaghe kelationRecord(items=frozenset(('ground beef', 'spaghetti', 'grated cheese')), support=0.005333333333333, ordered_statistics=[Orderedstatistic(items_base=frozenset(('spaghetti', 'grated cheese') telationRecord(items=frozenset({'ground beef', 'herb & pepper', 'mineral water'}), support=0.0066666666666666, ordered statistics=[OrderedStatistic(items base=frozenset({'mineral water', 'herb & RelationRecord(items=frozenset(('ground beef', 'nan', 'herb & pepper')), support=0.016, ordered_statistics=[OrderedStatistic(items_base=frozenset(('herb & pepper')), items_add=frozenset(('ground be RelationRecord(items=frozenset(('ground beef', 'spaghetti', 'herb & pepper'}), support=0.0064, ordered statistics=[OrderedStatistic(items base=frozenset(('ground beef', 'spaghetti', 'herb & pepper'}), items add=f telationRecord(items-frozenset({'ground beef', 'milk', 'olive oil'}), support=0.0049333333333333, ordered_statistics=[OrderedStatistic(items_base=frozenset({'ground beef', 'milk')}, items_add=frozenset([OrderedStatistic]) RelationRecord(items=frozenset(('ground beef', 'nan', 'tomato sauce')), support=0.005333333333333, ordered statistics=[OrderedStatistic(items base=frozenset(('tomato sauce')), items add=frozens

```
[ ] print(association_results[0])

RelationRecord(items=frozenset({'chicken', 'light cream'}), support=0.004533333333333334,
```

12.4.5 Mencetak Rules, Support, Confidence, dan Lift Ratio

Kode berikut menampilkan aturan asosiasi, support, confidence, dan lift ratio untuk setiap aturan asosiasi:

```
[ ] for item in association_results:
    pair = item[0]
    items = [x for x in pair]

    print("Rule: " + items[0] + " -> " + items[1])
    print("Support: " + str(item[1]))
    print("Confidence: " + str(item[2][0][2]))
    print("Lift: " + str(item[2][0][3]))
    print("=========="")
```

Rule: chicken -> light cream Support: 0.0045333333333333334 Confidence: 0.2905982905982906 Lift: 4.843304843304844 Rule: mushroom cream sauce -> escalope Support: 0.0057333333333333333 Confidence: 0.30069930069930073 Lift: 3.7903273197390845 Rule: pasta -> escalope Support: 0.0058666666666666667 Confidence: 0.37288135593220345 Lift: 4.700185158809287 Rule: ground beef -> herb & pepper Support: 0.016 Confidence: 0.3234501347708895 Lift: 3.2915549671393096 Rule: ground beef -> tomato sauce Support: 0.0053333333333333333 Confidence: 0.37735849056603776 Lift: 3.840147461662528 Rule: whole wheat pasta -> olive oil Support: 0.008 Confidence: 0.2714932126696833 Lift: 4.130221288078346 Rule: shrimp -> pasta Lift: 4.514493901473151 Rule: nan -> chicken Support: 0.004533333333333333 Confidence: 0.2905982905982906 Lift: 4.843304843304844 Rule: shrimp -> frozen vegetables Support: 0.0053333333333333333 Confidence: 0.23255813953488372 Lift: 3.260160834601174 Rule: ground beef -> cooking oil Support: 0.0048 Confidence: 0.5714285714285714 Lift: 3.281557646029315

Nilai support untuk aturan pertama adalah 0.0045. Jumlah ini dihitung dengan membagi jumlah transaksi yang mengandung light cream dibagi dengan jumlah total transaksi. Tingkat confidence untuk aturan tersebut adalah 0.2905 yang menunjukkan bahwa dari semua transaksi yang mengandung light cream, 29.05% transaksi juga mengandung chicken. Terakhir, lift 4.84 berarti bahwa pembelian chicken 4.84 kali lebih mungkin dibeli oleh pelanggan yang membeli light cream dibandingkan dengan kemungkinan penjualan ayam secara umum.

TUGAS

Terdapat dataset pada Grocery Store Dataset yang dapat diunduh pada link berikut: https://gitea.ums.ac.id/yusufsn/Praktikum_DWDM/src/branch/master/Data/Bab12/GroceryStoreDataSet.csv

Pada dataset tersebut terdapat 19 data transaksi dengan daftar item sebagai berikut:

	product
0	MILK,BREAD,BISCUIT
1	BREAD,MILK,BISCUIT,CORNFLAKES
2	BREAD,TEA,BOURNVITA
3	JAM,MAGGI,BREAD,MILK
4	MAGGI,TEA,BISCUIT
5	BREAD,TEA,BOURNVITA
6	MAGGI,TEA,CORNFLAKES
7	MAGGI,BREAD,TEA,BISCUIT
8	JAM,MAGGI,BREAD,TEA
9	BREAD,MILK
10	COFFEE,COCK,BISCUIT,CORNFLAKES
11	COFFEE,COCK,BISCUIT,CORNFLAKES
12	COFFEE,SUGER,BOURNVITA
13	BREAD,COFFEE,COCK
14	BREAD,SUGER,BISCUIT
15	COFFEE,SUGER,CORNFLAKES
16	BREAD,SUGER,BOURNVITA
17	BREAD,COFFEE,SUGER
18	BREAD,COFFEE,SUGER
19	TEA,MILK,COFFEE,CORNFLAKES

Kemudian kerjakanlah soal-soal berikut ini:

```
#1
GroceryStore = pd.read_csv("/content/drive/MyDrive/dwdm/GroceryStoreDataSet.csv", names = ["product"], header=None)

[ ] records = list(GroceryStore["product"].apply(lambda x: x.split(",")))
    print(records)

[ ['MILK', 'BREAD', 'BISCUIT'], ['BREAD', 'MILK', 'BISCUIT', 'CORNFLAKES'],
```

1. Dengan ketentuan jumlah minimum support = 0.3% dan minumum confidence = 20%. Tuliskan hasil aturan asosiasi yang dihasilkan?

RelationRecord(items_frozenset(('BISCUIT', 'COCK', 'CORNFLAKES')), support=0.1, ordered_statistics=[OrderedStatistic(items_base=frozenset(('COCK')), items_add=frozenset(('BISCUIT', 'CORNFLAKES')), or RelationRecord(items=frozenset({'BISCUIT', 'COFFEE', 'CORNFLAKES'}), support=0.1, ordered statistics=[OrderedStatistic(items base=frozenset({'CORNFLAKES'}), items add=frozenset({'BISCUIT', 'COFFEE' elationRecord(items=frozenset({'BISCUIT', 'TEA', 'MAGGI'}), support=0.1, ordered_statistics=[OrderedStatistic(items_base=frozenset({'MAGGI'}), items_add=frozenset({'BISCUIT', 'TEA'}), confidence=0 Record(items_frozenset({'BREAD', 'CORNFLAKES', 'MILK'}), support=0.05, ordered_statistic(=[OrderedStatistic(items_base=frozenset({'MILK'}), items_add-frozenset({'BREAD', 'CORNFLAKES'}), co RelationRecord(items=frozenset({'MAGGI', 'BREAD', '3AM'}), support=0.1, ordered statistics=[OrderedStatistic(items base=frozenset({'3AM'}), items add=frozenset({'MAGGI', 'BREAD'}), confidence=1.0, telationRecord(items_frozenset(('CORNFLAKES', 'COFFEE', 'COCK')), support=0.1, ordered_statistics=[OrderedStatistic(items_base=frozenset(('COCK')), items_add=frozenset(('COFFEE', 'CORNFLAKES')), ci Record(items=frozenset({'MILK', 'COFFEE', 'CORNFLAKES'})), support=0.05, ordered_statistics=[Orderedstatistic(items_base=frozenset({'MILK', 'COFFEE'}), items_add=frozenset({'CORNFLAKES'}), c RelationRecord(items=frozenset({'TEA', 'COFFEE', 'CORNFLAKES'}), support=0.05, ordered statistics=[OrderedStatistic(items base=frozenset({'TEA', 'COFFEE'}), items add=frozenset({'CORNFLAKES'}), con elationRecord(items=frozenset({'TEA', 'MILK', 'COFFEE'}), support=0.05, ordered_statistics=[Orderedstatistic(items_base=frozenset({'MILK'}), items_add=frozenset({'TEA', 'COFFEE'}), confidence=0.2, RelationRecord(items=frozenset({'MAGGI', 'MILK', 'JAM'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset({'AM'}), items_add=frozenset({'MAGGI', 'MILK')}, confidence=0.5, li RelationRecord(items=frozenset({'MAGGI', 'TEA', 'JAM'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset({'MAGGI'}), items_add=frozenset({'TEA', 'JAM'}), confidence=0.2, lift RelationRecord(items=frozenset({'BISCUIT', 'TEA', 'BREAD', 'MAGGI'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset({'MAGGI'}), items_add=frozenset({'BISCUIT', 'TEA', 'BREAD', RelationRecord(items=frozenset({'BISCUIT', 'COFFEE', 'COCK', 'CORNFLAKES')), support=0.1, ordered statistics=[OrderedStatistic(items base=frozenset({'COCK'}), items add=frozenset({'BISCUIT', 'COFFEE elationRecord(items=frozenset(('MAGGI', 'MILK', 'BREAD', 'JAM')), support=0.85, ordered_statistics[OrderedStatistic(items_base=frozenset(('JAM')), items_add=frozenset(('MAGGI', 'MILK', 'BREAD')), elationRecord(items=frozenset({'MAGGI', 'TEA', 'BREAD', 'JAM'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset(('JAM')), items_add=frozenset({'MAGGI', 'TEA', 'BREAD'}), c RelationRecord(items=frozenset({'TEA', 'MILK', 'COFFEE', 'CORNFLAKES'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset({'MILK'}), items_add=frozenset({'TEA', 'COFFEE', 'CO

```
Confidence: 1.0
Lift: 3.3333333333333333
Rule: TEA -> COFFEE
                                                                                   To exit full scree
Support: 0.05
Confidence: 1.0
Lift: 3.3333333333333333
Rule: TEA -> MILK
Support: 0.05
Confidence: 0.2
Lift: 4.0
       _____
Rule: TEA -> MILK
Support: 0.05
Confidence: 1.0
Lift: 3.33333333333333333333
Rule: MAGGI -> MILK
Support: 0.05
Confidence: 0.5
Lift: 10.0
Rule: MAGGI -> TEA
Support: 0.05
Confidence: 0.2
Lift: 4.0
Rule: BISCUIT -> BREAD
Support: 0.05
Confidence: 0.2
Lift: 4.0
Rule: BISCUIT -> TEA
Support: 0.05
Confidence: 0.2
Lift: 4.0
Rule: BISCUIT -> COFFEE
Support: 0.1
Confidence: 0.6666666666666666667
Lift: 6.66666666666667
       Rule: MAGGI -> MILK
Support: 0.05
Confidence: 0.5
Lift: 10.0
Rule: MAGGI -> TEA
Support: 0.05
```

2. Dengan ketentuan jumlah minimum support = 0.7% dan minumum confidence = 60%. Tuliskan hasil aturan asosiasi yang dihasilkan?

elationRecord(items=frozenset({'MAGGI', 'JAM'}), support=0.1, ordered_statistics=[OrderedStatistic(items_base=frozenset({'JAM'}), items_add=frozenset({'MAGGI'}), confidence=1.0, lift=4.0)]) lationRecord(items=frozenset(('BISCUIT', 'COFFEE', 'COCK')), support=0.1, ordered_statistics=[orderedstatistic(items_base=frozenset(('COCK')), items_add=frozenset(('BISCUIT', 'COFFEE')), confi ccord(items=frozenset(('BISCUIT', 'TEA', 'MAGGI')), support=0.1, ordered_statistics=(Ordered/statistic(items_base=frozenset(('BISCUIT', 'TEA')), items_add=frozenset(('MAGGI')), confidence= cord(items-frozenset({'SUGER', 'COFFEE', 'BOURNVITA'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset({'COFFEE', 'BOURNVITA'}), items_add=frozenset({'SUGER', 'COFFEE', 'COFFEE', 'BOURNVITA'}), items_add=frozenset({'SUGER', 'COFFEE', 'BOURNVITA'}), items_add=frozenset({'SUGER', 'COFFEE', 'COFFEE', 'BOURNVITA'}), items_add=frozenset({'SUGER', 'COFFEE', 'COFFEE cord(items=frozenset({'BREAD', 'CORNFLAKES', 'MILK'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset({'BREAD', 'CORNFLAKES'}), items_add=frozenset({'MILK'}), cor ord(items=frozenset(('MAGGI', 'BREAD', 'JAM')), support=0.1, ordered_statistics=[OrderedStatistic(items_base=frozenset(('JAM')), items_add=frozenset(('MAGGI', 'BREAD')), confidence=1.0, rd(items=frozenset({'CORNIFLAKES', 'COFFEE', 'COCK'})), support=0.1, ordered_statistics=[OrderedStatistic(items_base=frozenset({'COCK'})), items_add=frozenset(('COFFEE', 'CORNIFLAKES')), c ord(items-frozenset(('TEA', 'MILK', 'COFFEE')), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset(('TEA', 'COFFEE')), items_add=frozenset(('MILK')), confidence=1.0, cord(items-frozenset(('TEA', 'MILK', 'CORNFLAKES')), support-0.05, ordered_statistics-[Orderedstatistic(items_base-frozenset(('TEA', 'MILK')), items_add-frozenset(('CORNFLAKES')), confid ecord(items=frozenset({'MAGGI', 'MILK', 'JAM'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset({'MILK', 'JAM'}), items_add=frozenset({'MAGGI'}), confidence=1.0, li ecord(items=frozenset({'MAGGI', 'TEA', 'JAM'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset({'TEA', 'JAM'}), items_add=frozenset({'MAGGI'}), confidence=1.0, lif cord(items=frozenset({'BISCUIT', 'BREAD', 'CORNFLAKES', 'MILK'}), support=0.05, ordered statistics=[OrderedStatistic(items base=frozenset({'BREAD', 'CORNFLAKES'}), items add=fr cord(items=frozenset({'BISCUIT', 'TEA', 'BREAD', 'MAGGI'}), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset({'BISCUIT', 'TEA', 'BREAD'}), items_add=froz elationRecord(items-frozenset(('BISCUIT', 'COFFE', 'COCK', 'CORNFLAKES')), support=0.1, ordered_statistics=[OrderedStatistic(items_base=frozenset(('COCK')), items_add=frozenset(('BISCUIT', 'COFFE elationRecord(items=frozenset(('TEA', 'MILK', 'COFFEE', 'CORNFLAKES')), support=0.05, ordered_statistics=[OrderedStatistic(items_base=frozenset(('MILK', 'COFFEE')), items_add=frozenset(('TEA', 'C

print(association_results[0])



RelationRecord(items=frozenset({'MAGGI', 'JAM'}), support=0.1, ordered_statistic

```
[ ] for item in association_results:
       pair = item[0]
       items = [x for x in pair]
       print("Rule: " + items[0] + " -> " + items[1])
       print("Support: " + str(item[1]))
       print("Confidence: " + str(item[2][0][2]))
       print("Lift: " + str(item[2][0][3]))
       print("======="")
```

```
Rule: MAGGI -> JAM
Support: 0.1
Confidence: 1.0
Rule: BISCUIT -> COFFEE
Support: 0.1
Confidence: 0.666666666666667
Lift: 6.66666666666667
Rule: BISCUIT -> COCK
Support: 0.1
Confidence: 0.6666666666666667
Lift: 4.44444444444455
Rule: BISCUIT -> COFFEE
Support: 0.1
Confidence: 1.0
Lift: 3.333333333333333333
Rule: BISCUIT -> TEA
Support: 0.1
Confidence: 1.0
Lift: 4.0
Rule: SUGER -> COFFEE
Support: 0.05
Confidence: 1.0
Lift: 3.33333333333333333
Rule: BREAD -> CORNFLAKES
Support: 0.05
Confidence: 1.0
Lift: 4.0
Rule: MAGGI -> BREAD
Support: 0.1
Confidence: 1.0
Lift: 6.666666666666666
Rule: CORNFLAKES -> COFFEE
Support: 0.1
Confidence: 0.666666666666667
Lift: 3.333333333333333
Rule: MILK -> COFFEE
Support: 0.05
Confidence: 1.0
Lift: 3.33333333333333333
```

3. Jelaskan bagaimana nilai minimum support dan minumum confidence memengaruhi aturan asosiasi dan nilai lift ratio yang dihasilkan.

Penetapan nilai minimum support dan minimum confidence harus seimbang agar aturan yang dihasilkan tidak terlalu banyak (*overfitting*) atau terlalu sedikit (*underfitting*).