

## ▼ Knowledge Discovery & Data Mining Lab-05

Name: Gurvinder Kaur Matharu

PRN: 20190802077

### ▼ AIM:

To implement the Apriori Algorithm using WEKA and python.

### ▼ Apriori Algorithm using WEKA

Dataset used:

	A	B	C	D	E	F	G
1	Transaction ID	BREAD	BUTTER	JAM	SUGAR	MILK	COOKIES
2	1	YES		YES	YES		
3	2			YES		YES	YES
4	3	YES			YES	YES	
5	4	YES	YES				YES
6	5		YES		YES	YES	YES

The screenshot shows the Weka Explorer application window. The 'Preprocess' tab is active. The 'Filter' section shows 'None' selected. The 'Current relation' section displays 'Relation: Dataset-weka filters unsupervised attribute.Remove-R1' and 'Instances: 5'. The 'Attributes' section shows a list of attributes: BREAD, BUTTER, JAM, SUGAR, MILK, and COOKIES. The 'Selected attribute' section shows 'Name: BREAD', 'Missing: 2 (40%)', 'Distinct: 1', and 'Type: Nominal'. A table below shows the selected attribute data: 'No. 1', 'Label YES', 'Count 3', and 'Weight 3.0'. The 'Class: COOKIES (Nom)' is selected at the bottom. A visualization of the data is shown as a bar chart with a blue bar for 'YES' and a black bar for 'NO'.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -I -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1

Start Stop

Result list (right-click...)

- 13:19:55 - Apriori
- 13:20:22 - Apriori
- 13:31:45 - Apriori
- 13:35:42 - Apriori
- 13:36:11 - Apriori
- 13:36:36 - Apriori
- 13:57:44 - Apriori
- 13:58:01 - Apriori
- 17:11:12 - Apriori
- 17:11:29 - Apriori

Associator output

```
=== Run information ===

Scheme:      weka.associations.Apriori -I -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1
Relation:    Dataset-weka.filters.unsupervised.attribute.Remove-RI
Instances:   5
Attributes:  6
             BREAD
             BUTTER
             JAM
             SUGAR
             MILK
             COOKIES

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

Apriori
=====

Minimum support: 0.3 (1 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Large Itemsets L(1):
BREAD=YES 3
BUTTER=YES 2
JAM=YES 2
SUGAR=YES 3
MILK=YES 3
COOKIES=YES 3
```

Status

OK Log x 0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -I -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1

Start Stop

Result list (right-click...)

- 13:19:55 - Apriori
- 13:20:22 - Apriori
- 13:31:45 - Apriori
- 13:35:42 - Apriori
- 13:36:11 - Apriori
- 13:36:36 - Apriori
- 13:57:44 - Apriori
- 13:58:01 - Apriori
- 17:11:12 - Apriori
- 17:11:29 - Apriori

Associator output

```
Size of set of large itemsets L(2): 14

Large Itemsets L(2):
BREAD=YES BUTTER=YES 1
BREAD=YES JAM=YES 1
BREAD=YES SUGAR=YES 2
BREAD=YES MILK=YES 1
BREAD=YES COOKIES=YES 1
BUTTER=YES SUGAR=YES 1
BUTTER=YES MILK=YES 1
BUTTER=YES COOKIES=YES 2
JAM=YES SUGAR=YES 1
JAM=YES MILK=YES 1
JAM=YES COOKIES=YES 1
SUGAR=YES MILK=YES 2
SUGAR=YES COOKIES=YES 1
MILK=YES COOKIES=YES 2

Size of set of large itemsets L(3): 8

Large Itemsets L(3):
BREAD=YES BUTTER=YES COOKIES=YES 1
BREAD=YES JAM=YES SUGAR=YES 1
BREAD=YES SUGAR=YES MILK=YES 1
BUTTER=YES SUGAR=YES MILK=YES 1
BUTTER=YES SUGAR=YES COOKIES=YES 1
BUTTER=YES MILK=YES COOKIES=YES 1
JAM=YES MILK=YES COOKIES=YES 1
SUGAR=YES MILK=YES COOKIES=YES 1

Size of set of large itemsets L(4): 1

Large Itemsets L(4):
```

Status

OK Log x 0



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -I -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c 1

Start Stop

Result list (right-click to open):

- 13:19:55 - Apriori
- 13:20:22 - Apriori
- 13:31:45 - Apriori
- 13:35:42 - Apriori
- 13:36:11 - Apriori
- 13:36:36 - Apriori
- 13:57:44 - Apriori
- 13:58:01 - Apriori
- 17:11:12 - Apriori
- 17:11:29 - Apriori

Associator output

```
party snack foods=t 2330
tissues-paper prd=t 2247
soft drinks=t 1888
cheese=t 1879
milk-cream=t 2939
margarine=t 2288
dairy foods=t 1669
beef=t 1739
fruit=t 2962
vegetables=t 2961
stationary=t 1457
total=low 2948
total=high 1679

Size of set of large itemsets L(2): 68

Large Itemsets L(2):
bread and cake=t baking needs=t 2191
bread and cake=t juice-sat-cord-ms=t 1869
bread and cake=t biscuits=t 2083
bread and cake=t breakfast food=t 1441
bread and cake=t sauces-gravy-pkle=t 1710
bread and cake=t frozen foods=t 2129
bread and cake=t pet foods=t 1490
bread and cake=t party snack foods=t 1808
bread and cake=t tissues-paper prd=t 1776
bread and cake=t soft drinks=t 1429
bread and cake=t cheese=t 1492
bread and cake=t milk-cream=t 2337
bread and cake=t margarine=t 1831
bread and cake=t fruit=t 2325
bread and cake=t vegetables=t 2298
bread and cake=t total=low 1917
```

Status

OK Log x 0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -I -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c 1

Start Stop

Result list (right-click to open):

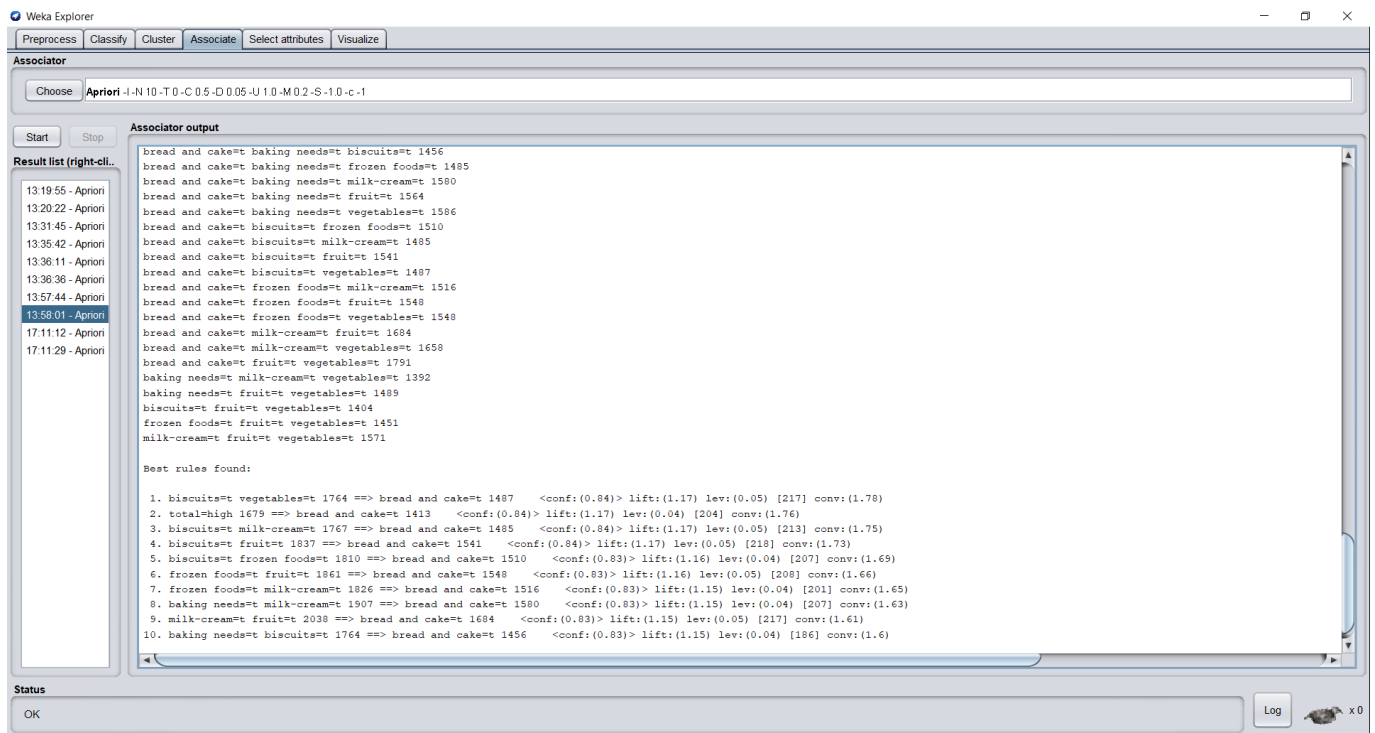
- 13:19:55 - Apriori
- 13:20:22 - Apriori
- 13:31:45 - Apriori
- 13:35:42 - Apriori
- 13:36:11 - Apriori
- 13:36:36 - Apriori
- 13:57:44 - Apriori
- 13:58:01 - Apriori
- 17:11:12 - Apriori
- 17:11:29 - Apriori

Associator output

```
bread and cake=t total=high 1413
baking needs=t juice-sat-cord-ms=t 1619
baking needs=t biscuits=t 1764
baking needs=t sauces-gravy-pkle=t 1521
baking needs=t frozen foods=t 1835
baking needs=t party snack foods=t 1530
baking needs=t tissues-paper prd=t 1573
baking needs=t milk-cream=t 1907
baking needs=t margarine=t 1645
baking needs=t fruit=t 1900
baking needs=t vegetables=t 1949
baking needs=t total=low 1515
juice-sat-cord-ms=t biscuits=t 1542
juice-sat-cord-ms=t frozen foods=t 1619
juice-sat-cord-ms=t party snack foods=t 1482
juice-sat-cord-ms=t milk-cream=t 1637
juice-sat-cord-ms=t fruit=t 1672
juice-sat-cord-ms=t vegetables=t 1658
biscuits=t sauces-gravy-pkle=t 1406
biscuits=t frozen foods=t 1810
biscuits=t party snack foods=t 1592
biscuits=t tissues-paper prd=t 1453
biscuits=t milk-cream=t 1767
biscuits=t margarine=t 1493
biscuits=t fruit=t 1837
biscuits=t vegetables=t 1764
sauces-gravy-pkle=t frozen foods=t 1497
sauces-gravy-pkle=t milk-cream=t 1487
sauces-gravy-pkle=t fruit=t 1490
sauces-gravy-pkle=t vegetables=t 1558
frozen foods=t party snack foods=t 1582
frozen foods=t tissues-paper prd=t 1505
frozen foods=t milk-cream=t 1826
```

Status

OK Log x 0



## ▼ Apriori Algorithm using python

```
!pip install apyori
```

Requirement already satisfied: apyori in /usr/local/lib/python3.7/dist-packages (1.1.

```
import pandas as pd
from apyori import apriori
```

```
data = pd.read_csv('Market_Basket_Optimisation.csv', header=None)
```

```
data.head()
```

	0	1	2	3	4	5	6	7	8	9
0	shrimp	almonds	avocado	vegetables mix	green grapes	whole wheat flour	yams	cottage cheese	energy drink	tomato juice
1	burgers	meatballs	eggs	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	chutney	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	turkey	avocado	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
data.shape
```

```
(7501, 20)
```

```
transactions=[]
for i in range(0,7501):
    transactions.append([str(data.values[i,j]) for j in range(0,20)])
```

```
association_rules = apriori(transactions, min_support=0.005, min_confidence=0.2,min_lift =
association_results=list(association_rules)
```

```
print(len(association_results))
```

38

```
association_results
```

```
[< RelationRecord(items=frozenset({'escalope', 'mushroom cream sauce'}), support=0.0057
RelationRecord(items=frozenset({'escalope', 'pasta'}), support=0.005865884548726837,
RelationRecord(items=frozenset({'herb & pepper', 'ground beef'}), support=0.01599786
RelationRecord(items=frozenset({'tomato sauce', 'ground beef'}), support=0.005332622
RelationRecord(items=frozenset({'whole wheat pasta', 'olive oil'}), support=0.007998
RelationRecord(items=frozenset({'shrimp', 'pasta'}), support=0.005065991201173177, o
RelationRecord(items=frozenset({'shrimp', 'chocolate', 'frozen vegetables'}), suppor
RelationRecord(items=frozenset({'escalope', 'nan', 'mushroom cream sauce'}), support
RelationRecord(items=frozenset({'pasta', 'escalope', 'nan'}), support=0.005865884548
RelationRecord(items=frozenset({'spaghetti', 'ground beef', 'frozen vegetables'}), s
RelationRecord(items=frozenset({'shrimp', 'mineral water', 'frozen vegetables'}), su
RelationRecord(items=frozenset({'spaghetti', 'olive oil', 'frozen vegetables'}), sup
RelationRecord(items=frozenset({'shrimp', 'spaghetti', 'frozen vegetables'}), suppor
RelationRecord(items=frozenset({'spaghetti', 'tomatoes', 'frozen vegetables'}), supp
RelationRecord(items=frozenset({'spaghetti', 'grated cheese', 'ground beef'}), suppo
RelationRecord(items=frozenset({'herb & pepper', 'mineral water', 'ground beef'}), s
RelationRecord(items=frozenset({'herb & pepper', 'ground beef', 'nan'}), support=0.0
RelationRecord(items=frozenset({'herb & pepper', 'spaghetti', 'ground beef'}), suppo
RelationRecord(items=frozenset({'tomato sauce', 'ground beef', 'nan'}), support=0.00
RelationRecord(items=frozenset({'shrimp', 'spaghetti', 'ground beef'}), support=0.00
RelationRecord(items=frozenset({'milk', 'spaghetti', 'olive oil'}), support=0.007199
RelationRecord(items=frozenset({'mineral water', 'soup', 'olive oil'}), support=0.00
RelationRecord(items=frozenset({'whole wheat pasta', 'olive oil', 'nan'}), support=0
RelationRecord(items=frozenset({'pasta', 'nan', 'shrimp'}), support=0.00506599120117
RelationRecord(items=frozenset({'spaghetti', 'pancakes', 'olive oil'}), support=0.00
RelationRecord(items=frozenset({'shrimp', 'chocolate', 'frozen vegetables', 'nan'}),
RelationRecord(items=frozenset({'spaghetti', 'ground beef', 'frozen vegetables', 'na
RelationRecord(items=frozenset({'shrimp', 'mineral water', 'frozen vegetables', 'nan
RelationRecord(items=frozenset({'spaghetti', 'olive oil', 'frozen vegetables', 'nan'
RelationRecord(items=frozenset({'shrimp', 'spaghetti', 'frozen vegetables', 'nan'}),
RelationRecord(items=frozenset({'spaghetti', 'tomatoes', 'frozen vegetables', 'nan'})
RelationRecord(items=frozenset({'spaghetti', 'grated cheese', 'ground beef', 'nan'})
RelationRecord(items=frozenset({'herb & pepper', 'mineral water', 'ground beef', 'na
RelationRecord(items=frozenset({'herb & pepper', 'spaghetti', 'ground beef', 'nan'})
RelationRecord(items=frozenset({'shrimp', 'spaghetti', 'ground beef', 'nan'}), suppo
RelationRecord(items=frozenset({'milk', 'spaghetti', 'olive oil', 'nan'}), support=0
RelationRecord(items=frozenset({'mineral water', 'olive oil', 'soup', 'nan'}), suppo
RelationRecord(items=frozenset({'spaghetti', 'pancakes', 'olive oil', 'nan'}), suppo
```

```
results = pd.DataFrame(association_results)
results.head()
```

	items	support	ordered_statistics
0	(escalope, mushroom cream sauce)	0.005733	[((mushroom cream sauce), (escalope), 0.300699...
1	(escalope, pasta)	0.005866	[((pasta), (escalope), 0.3728813559322034, 4.7...
2	(herb & pepper, ground beef)	0.015998	[((herb & pepper), (ground beef), 0.3234501347...
3	(tomato sauce, ground beef)	0.005333	[((tomato sauce), (ground beef), 0.37735849056...
4	(whole wheat pasta, olive oil)	0.007999	[((whole wheat pasta), (olive oil), 0.27149321...