

Experiment 8: FP-Growth Algorithm

Aim: Generate Association Rules using the FP-Growth algorithm.

DESCRIPTION:

Association rule mining is a data mining technique used to find relationships between items in a dataset.

The FP-Growth (Frequent Pattern Growth) algorithm is an improved version of Apriori that finds frequent itemsets without generating candidate sets. It uses a compact data structure called the **FP-tree**, which stores item occurrences in a tree form and reduces database scans.

FP-Growth works in two main phases:

1. Building the FP-Tree from the dataset.
2. Extracting frequent patterns from the tree..

It is faster and more memory-efficient than Apriori, especially for large datasets.

ILLUSTRATION:

Build a decision tree for the following data

Milk	Bread	Butter	Jams	Eggs
Yes	Yes	Yes	No	No
Yes	Yes	No	No	No
No	Yes	Yes	No	No
Yes	No	Yes	Yes	No
	Yes	Yes	Yes	No
Yes	Yes	Yes	Yes	No
Yes	Yes	No	No	Yes
No	Yes	Yes	No	Yes
Yes	Yes	Yes	Yes	No

Minimum Support = 3 transactions

- 1) Item frequency:

Bread = 8

Butter = 7

Milk = 6

Jam = 2

Eggs = 2

- 2) (Remove items below support threshold: Jam and Eggs):

Keep frequent items {Bread, Butter, Milk}

Transactions used for FP-Tree building (with items sorted by frequency):

Bread = 8

Butter = 7

Milk = 6

Jam = 0

Eggs = 0

- 3) Frequent Itemsets found:

1-itemsets: {Bread}, {Butter}, {Milk}

2-itemsets: {Bread, Butter}, {Bread, Milk}, {Butter, Milk}

3-itemsets: {Bread, Butter, Milk}

- 4) Association Rules (min confidence = 70%)

{Bread, Butter} → {Milk} → Confidence = 75%

{Bread, Milk} → {Butter} → Confidence = 83%

{Butter, Milk} → {Bread} → Confidence = 88%

Strongest Rule: Customers buying Butter and Milk are most likely to buy Bread.

Procedure for running in weka:

Step 1: Creating the ARRF file

Open Notepad and type the following data exactly as shown below.

The screenshot shows a dark-themed text editor window. At the top, there's a header bar with a file icon, the filename "vats_1.arff", a close button ("X"), and a plus sign button ("+"). Below the header is a menu bar with "File", "Edit", and "View". The main content area contains the following ARFF code:

```
@relation groceries

@attribute Milk {yes, no}
@attribute Bread {yes, no}
@attribute Butter {yes, no}
@attribute Jam {yes, no}
@attribute Eggs {yes, no}

@data
yes,yes,yes,no,no
yes,yes,no,no,no
no,yes,yes,no,no
yes,no,yes,no,no
no,yes,yes,yes,no
yes,yes,yes,yes,no
yes,yes,no,no,yes
no,yes,yes,no,yes
yes,yes,yes,no,no
```

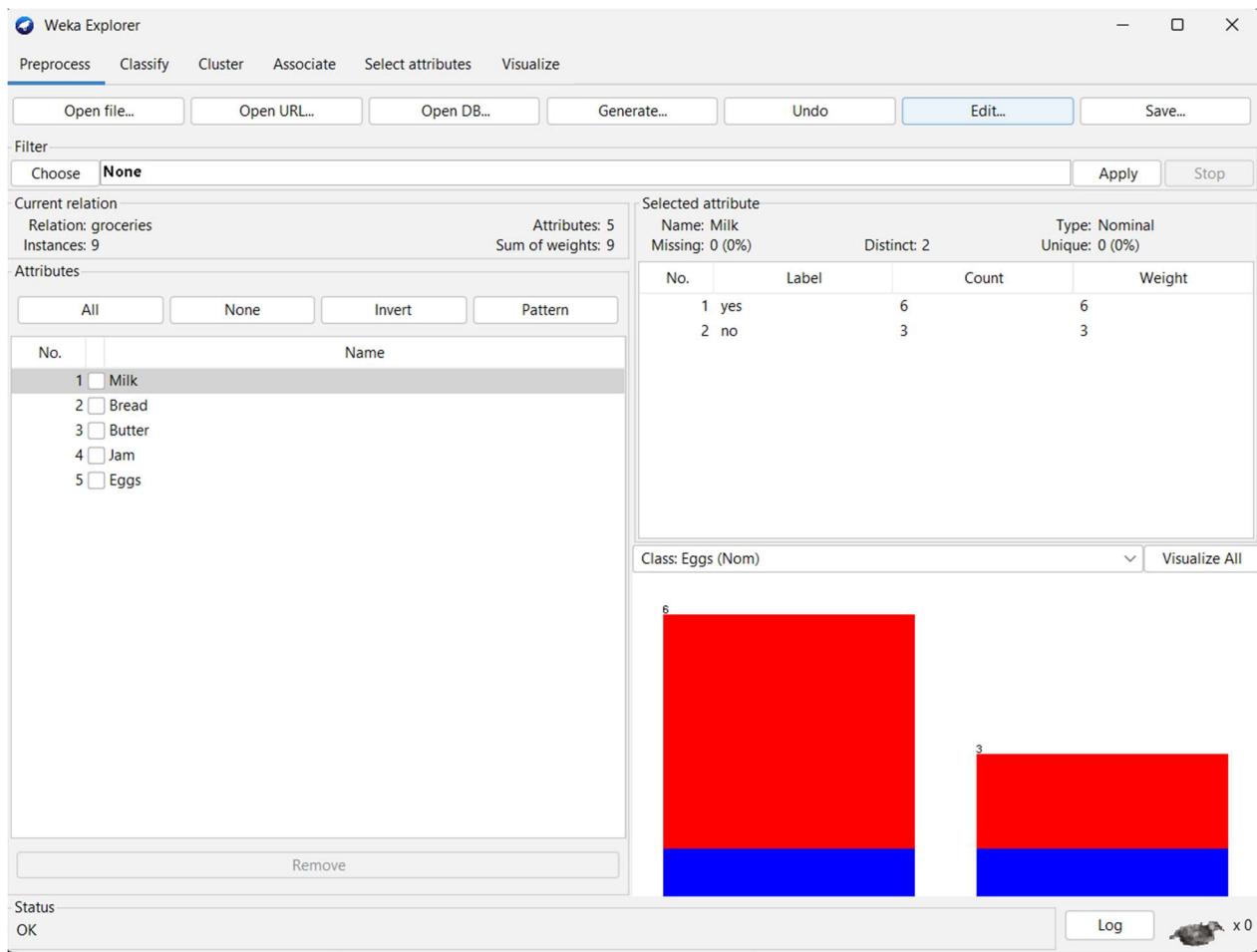
Save it with the following options:

File Type: *All Files*

Encoding: *UTF-8*

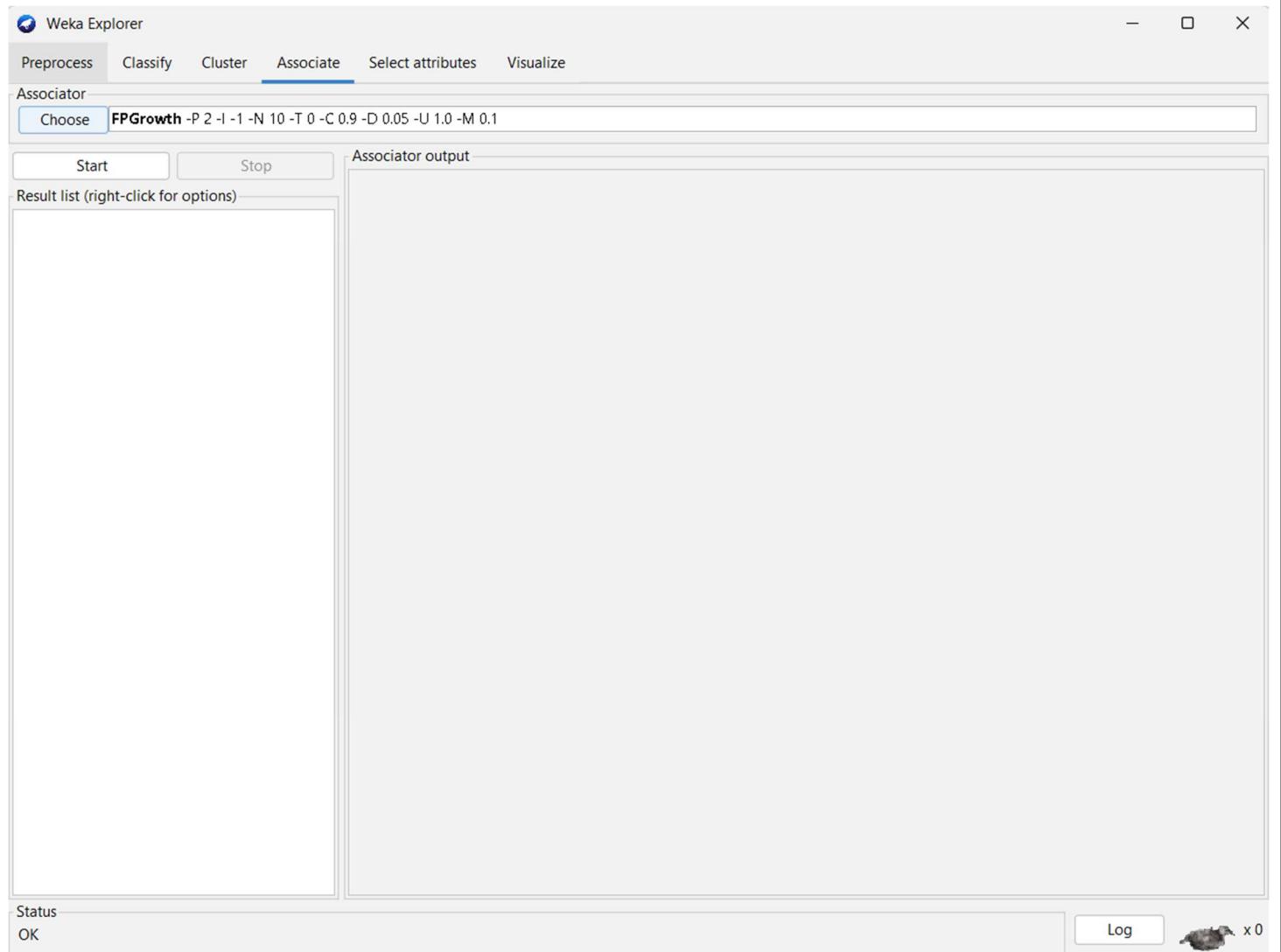
Step2:

Open **WEKA Explorer** → **Preprocess Tab** → **Open File**



Step 3:

Go to the **Associate Tab** → Choose algorithm **FP-Growth**



Step 4:

Click **Start**. WEKA will display frequent itemsets and generated association rules.

```

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Associate
Choose FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Start Stop
Result list (right-click for ...)
21:18:08 - FPGrowth
Associator output
==== Run information ====
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: groceries
Instances: 9
Attributes: 5
Milk
Bread
Butter
Jam
Eggs
==== Associator model (full training set) ====
FPGrowth found 7 rules (displaying top 7)

1. [Butter=no]: 2 ==> [Jam=no]: 2 <conf:(1)> lift:(1.29) lev:(0.05) conv:(0.44)
2. [Bread=no]: 1 ==> [Jam=no]: 1 <conf:(1)> lift:(1.29) lev:(0.02) conv:(0.22)
3. [Bread=no]: 1 ==> [Eggs=no]: 1 <conf:(1)> lift:(1.29) lev:(0.02) conv:(0.22)
4. [Eggs=no, Butter=no]: 1 ==> [Jam=no]: 1 <conf:(1)> lift:(1.29) lev:(0.02) conv:(0.22)
5. [Bread=no]: 1 ==> [Jam=no, Eggs=no]: 1 <conf:(1)> lift:(1.8) lev:(0.05) conv:(0.44)
6. [Jam=no, Bread=no]: 1 ==> [Eggs=no]: 1 <conf:(1)> lift:(1.29) lev:(0.02) conv:(0.22)
7. [Eggs=no, Bread=no]: 1 ==> [Jam=no]: 1 <conf:(1)> lift:(1.29) lev:(0.02) conv:(0.22)

```

Conclusion:

The FP-Growth algorithm efficiently discovers frequent itemsets without generating candidate sets. It is faster than Apriori and produces the same association results.

From the given dataset, strong rules identified include:

$$\begin{aligned} \{\text{Bread}, \text{Butter}\} &\rightarrow \{\text{Milk}\} \\ \{\text{Butter}, \text{Milk}\} &\rightarrow \{\text{Bread}\} \end{aligned}$$

FP-Growth is a practical method for market basket analysis and other association-based data mining tasks.

Exercise:

Create your own ARFF file with 6–8 transactions and at least five items. Run FP-Growth

