

Course Code and Name: CSI0803 Data Mining

PRACTICAL 9

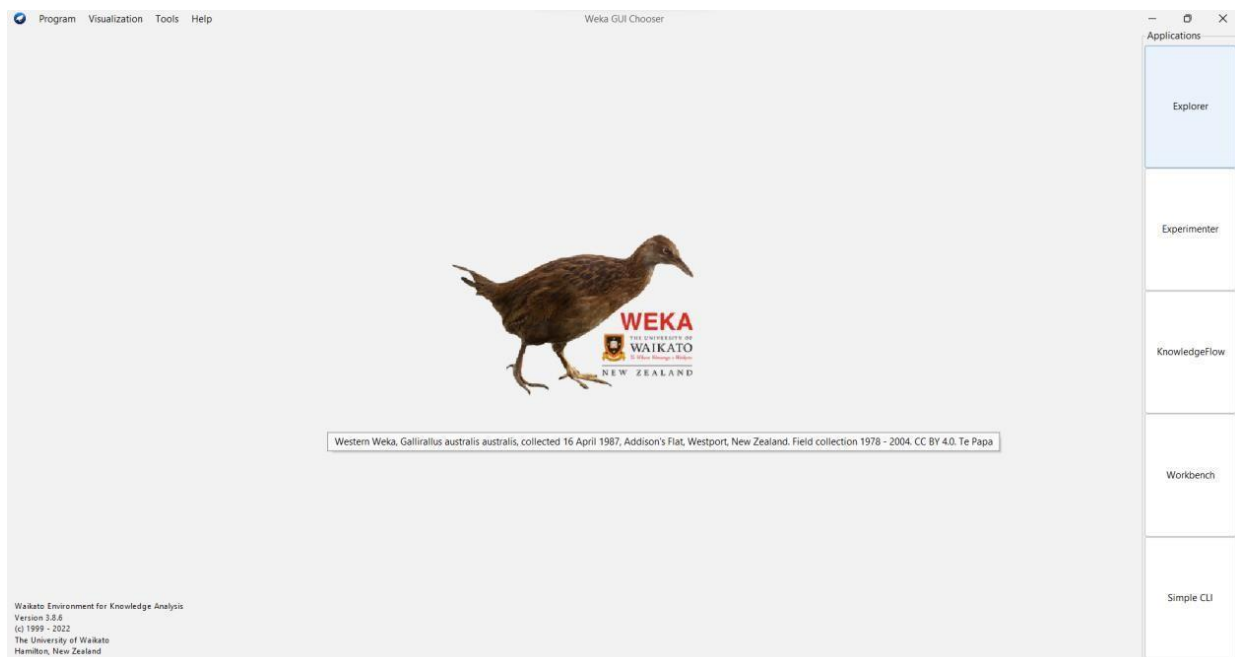
WEKA:

- WEKA is a collection of machine learning algorithms for data mining tasks. WEKA supports several data mining tasks, including data pre-processing, classification, regression, clustering, and association rule mining. It also includes a number of visualization and data exploration tools. WEKA provides a graphical user interface (GUI) as well as command-line interface for users to access its features. It can process data in various formats such as CSV, ARFF etc.
- WEKA also provides access to several readymade popular data sets for experimentation. It can be used for both supervised and unsupervised learning. It has several inbuilt machine learning algorithms, such as Random Forest, Naive Bayes, and Support Vector Machines, Decision trees etc. It also includes an evaluation module that allows users to compare the performance of different algorithms on a given dataset.
- WEKA can be integrated with other programming languages like Python, R through its API. WEKA's library can be used for developing new machine learning algorithms or modification of existing ones, as it is open-source.

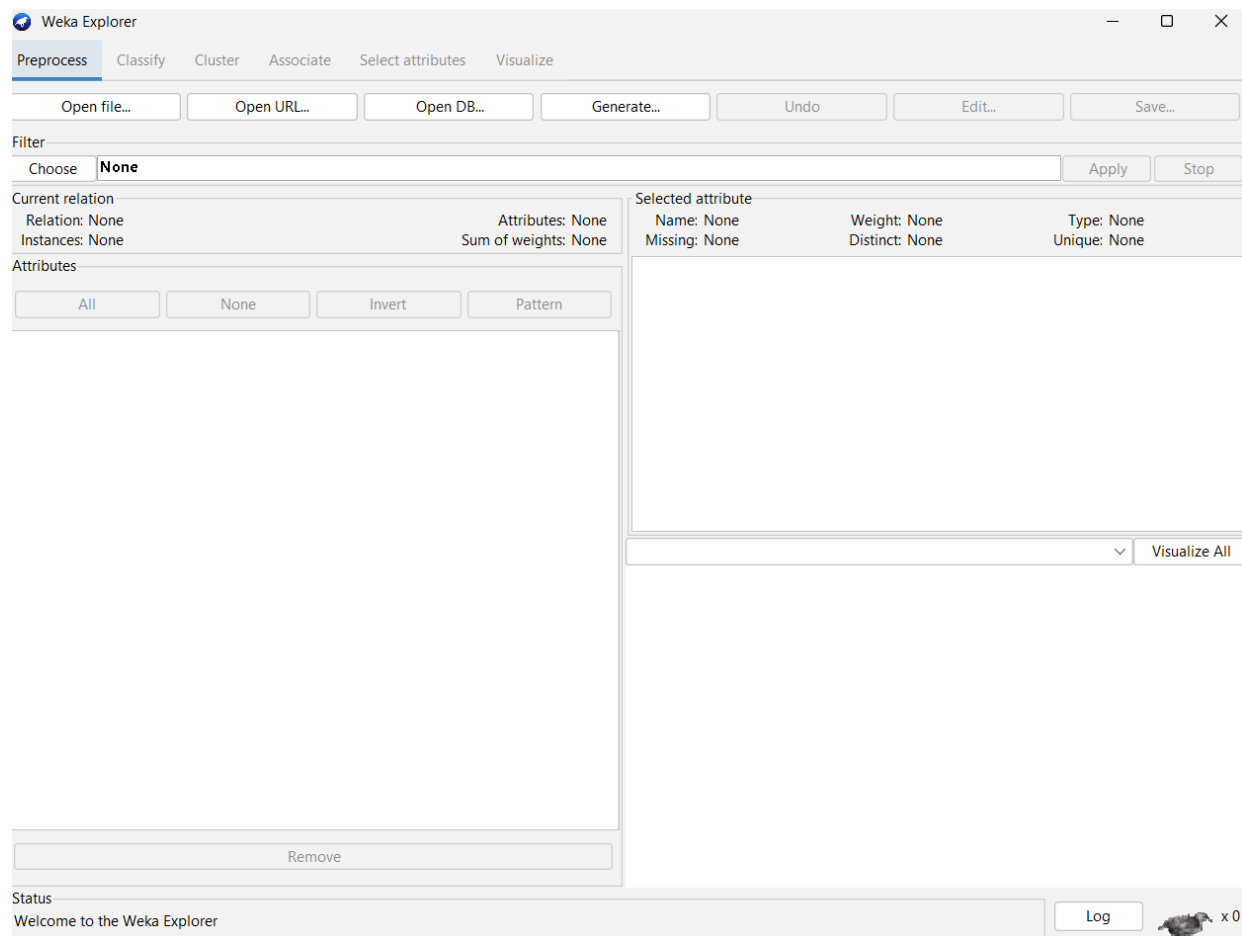
R-programming:

➤ R is a programming language and software environment for statistical computing and graphics. It is widely used among statisticians and data miners for data analysis and visualization. R has a large collection of packages and libraries for machine learning. Both WEKA and R can be used for machine learning tasks, but the syntax and usage are different. WEKA has a graphical user interface, while R is a programming language that is typically used in a command-line interface. R has a more extensive library for statistical and machine learning than WEKA.

Weka Homescreen:

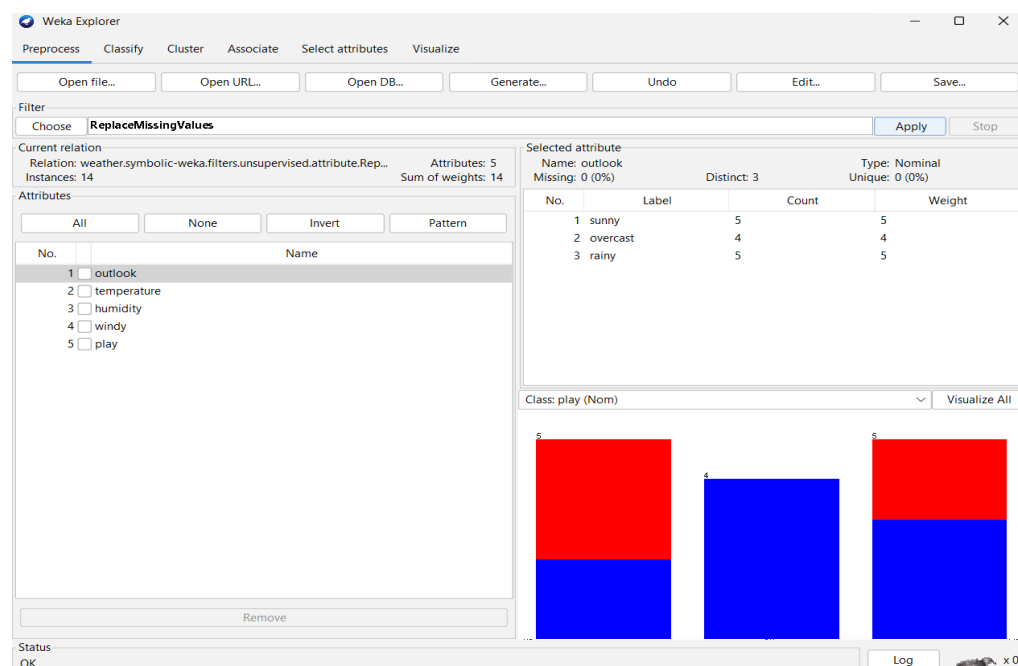


Opening explorer:



Trying on Supermarket dataset for getting frequent itemsets:

Replacing Missing Values:




After Preprocessing (Missing Percent – 0%):

Selected attribute			
Name: outlook		Type: Nominal	
Missing: 0 (0%)		Distinct: 3	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	sunny	5	5
2	overcast	4	4
3	rainy	5	5

Frequent Itemset Mining Using Apriori:

Setting Properties:

 weka.gui.GenericObjectEditor ✕

veka.associations.Apriori

About

Class implementing an Apriori-type algorithm. More Capabilities

car

False

▼

classIndex

-1

delta

0.05

doNotCheckCapabilities

False

▼

lowerBoundMinSupport

0.2

metricType

Confidence

▼

minMetric

0.9

numRules

10

outputItemSets

False

▼

removeAllMissingCols

False

▼

significanceLevel

-1.0

treatZeroAsMissing

False

▼

upperBoundMinSupport

1.0

verbose

False

▼

Open...

Save...

OK

Cancel

OUTPUT:

Associator output

```
=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1
Relation:    weather.symbolic-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.RemoveEmpty
Instances:   14
Attributes:  5
              outlook
              temperature
              humidity
              windy
              play
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.2 (3 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 26

Size of set of large itemsets L(3): 4
```

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
2. temperature=cool 4 ==> humidity=normal 4 <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
3. humidity=normal windy=FALSE 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
4. outlook=sunny play=no 3 ==> humidity=high 3 <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
5. outlook=sunny humidity=high 3 ==> play=no 3 <conf:(1)> lift:(2.8) lev:(0.14) [1] conv:(1.93)
6. outlook=rainy play=yes 3 ==> windy=FALSE 3 <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3 <conf:(1)> lift:(1.56) lev:(0.08) [1] conv:(1.07)
8. temperature=cool play=yes 3 ==> humidity=normal 3 <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)

Visualization:

