Name:	
Roll No:	
Class/Sem:	TE/V
Experiment No.:	4
Title:	Using open source tools Implement Classifiers
Date of Performance:	
Date of Submission:	
Marks:	
Sign of Faculty:	

Aim: To implement Naïve Bayes Classifier using open source tool WEKA.

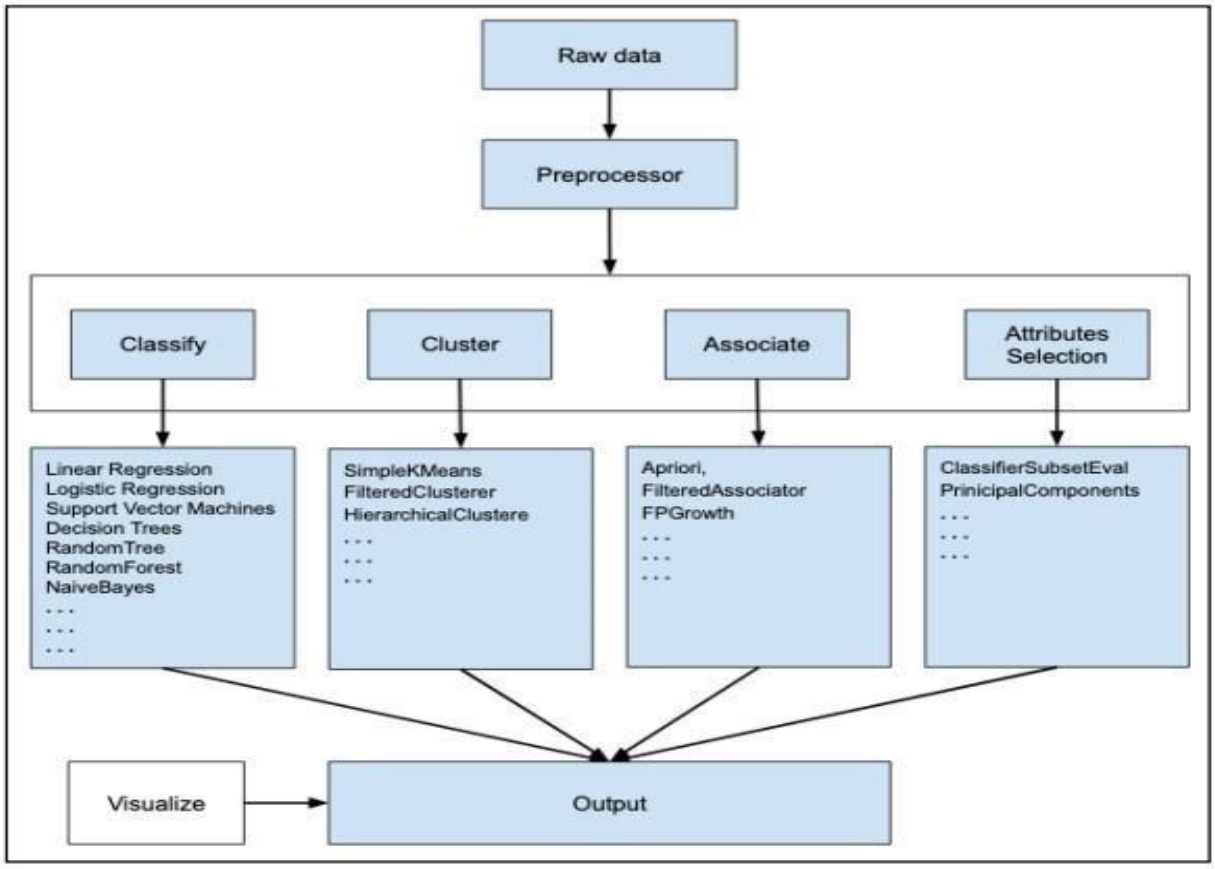
Objective: To make students well versed with open source tool like WEKA to implement Naïve Bayes Classifier.

Theory:

Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks.

WEKA:

WEKA – an open source software provides tools for data preprocessing, implementation of several data Mining algorithms, and visualization tools so that you can develop data mining techniques and apply them to real-world data mining problems. Weka is summarized in the following diagram:

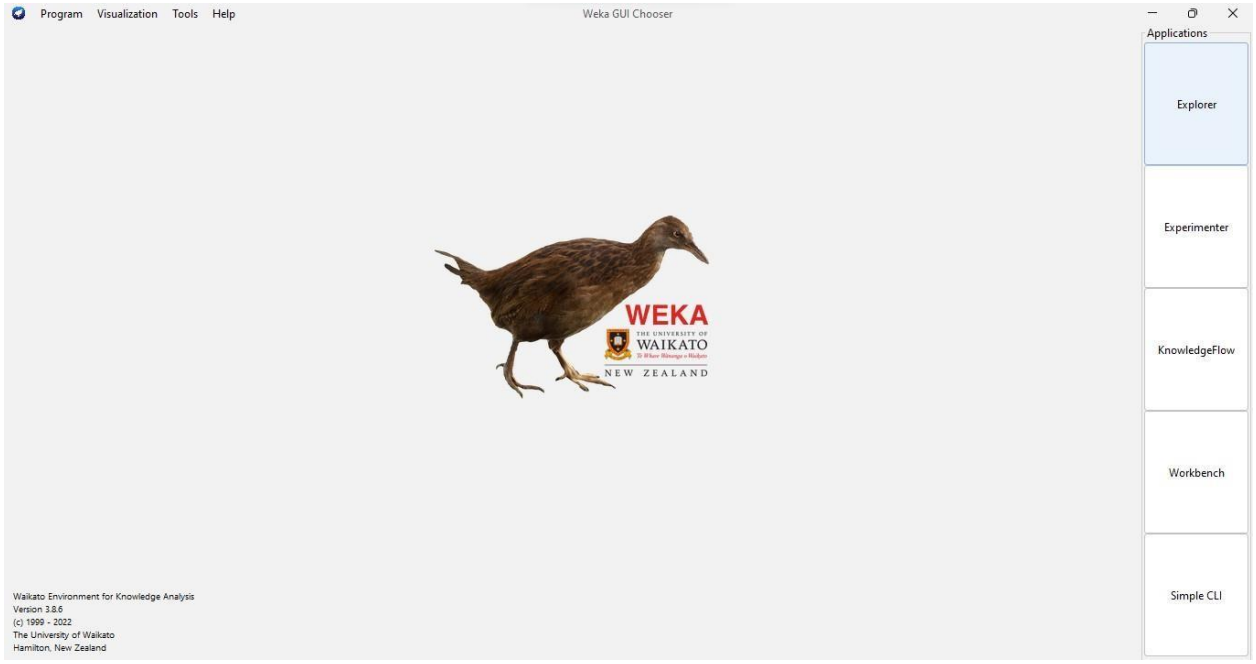


First, you will start with the raw data collected from the field. This data may contain several null values and irrelevant fields. You use the data preprocessing tools provided in WEKA to cleanse the data. Then, you would save the preprocessed data in your local storage for applying Data Mining algorithms.

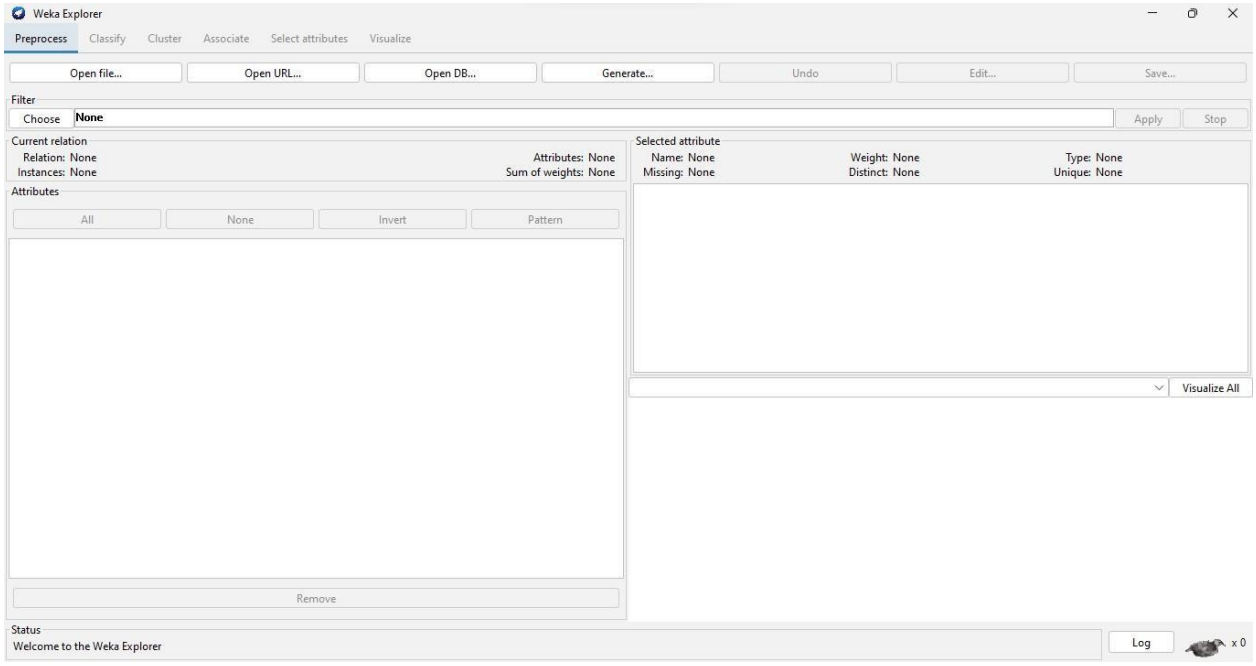
Next, depending on the kind of Data Mining model that you are trying to develop you would select one of the options such as Classify, Cluster, or Associate. The Attributes Selection allows the automatic selection of features to create a reduced dataset. Note that under each category, WEKA provides the implementation of several algorithms. You would select an algorithm of your choice, set the desired parameters and run it on the dataset. Then, WEKA would give you the statistical output of the model processing. It provides you a visualization tool to inspect the data. The various models can be applied on the same dataset. You can then compare the outputs of different models and select the best that meets your purpose.

Output:

Step 1:Open Weka,the following GUI should appear on your system

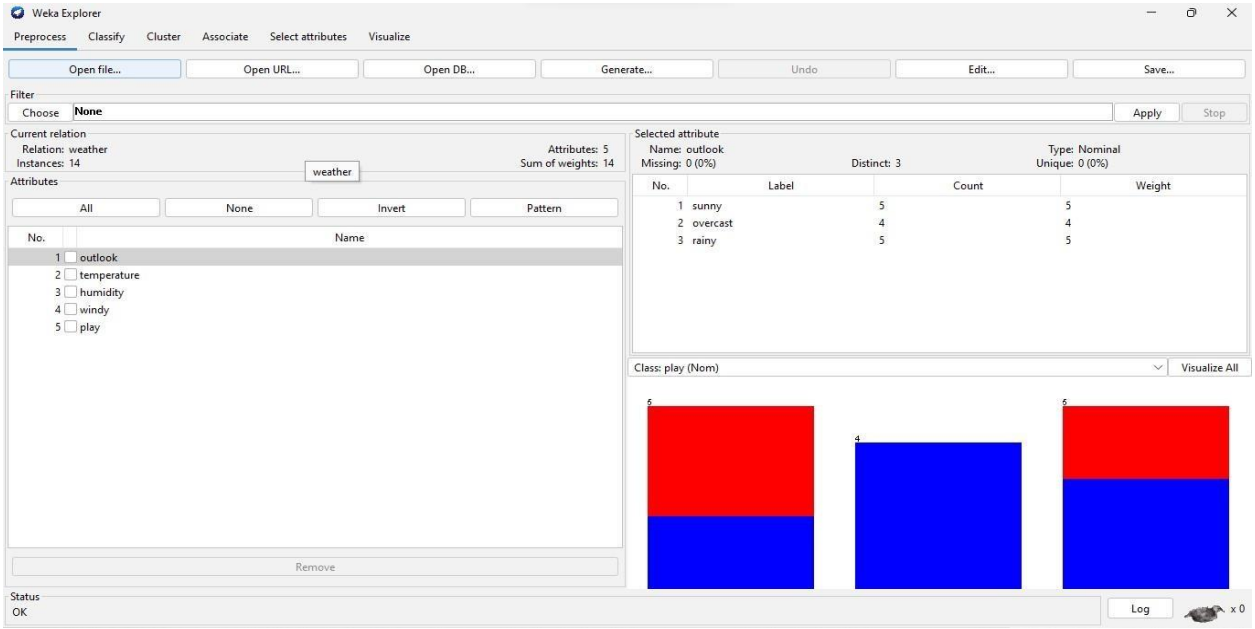


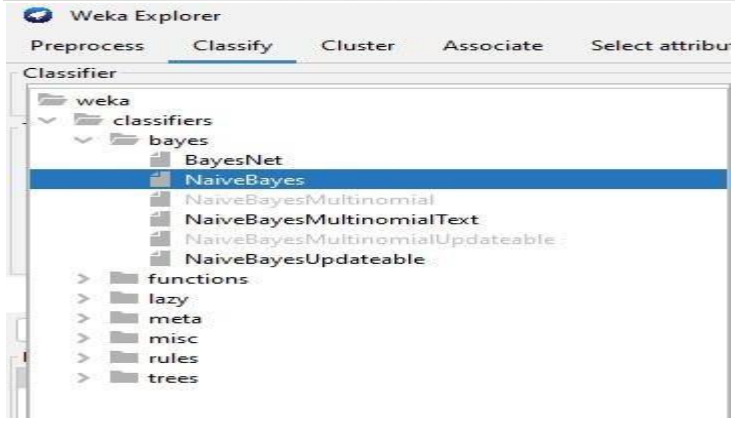
Step 2: Click on the Explorer.



Step 3:Select weather .numeric.arff data set from Open file section .The following screen would appear.

Step 4: Select Naive Bayes Classifier, under classifier -> Bayes ->Naive Bayes nad click on apply





Step 5: Click on start, the following output would appear.

Classifier output

```
   [total]          11.0     7.0



Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               64.2857 %
Incorrectly Classified Instances         5               35.7143 %
Kappa statistic                          0.1026
Mean absolute error                      0.4649
Root mean squared error                  0.543
Relative absolute error                 97.6254 %
Root relative squared error            110.051  %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.889    0.800    0.667      0.889    0.762      0.122    0.444     0.633     yes
                 0.200    0.111    0.500      0.200    0.286      0.122    0.444     0.397     no
Weighted Avg.    0.643    0.554    0.607      0.643    0.592      0.122    0.444     0.548

=== Confusion Matrix ===

 a b   <-- classified as
 8 1 | a = yes
 4 1 | b = no
```

Conclusion:

Implementing classifiers with open source tools is a practical and cost-effective way to harness the power of machine learning and data analysis. These tools enable you to build, train, and evaluate classifiers for a wide range of applications, from predictive analytics to image recognition and natural language processing. However, it's essential to choose the right tool for your specific problem and to stay engaged with the open source community for ongoing support and updates.