Name:	**Vinith Shetty**
Roll No:	**52**
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.

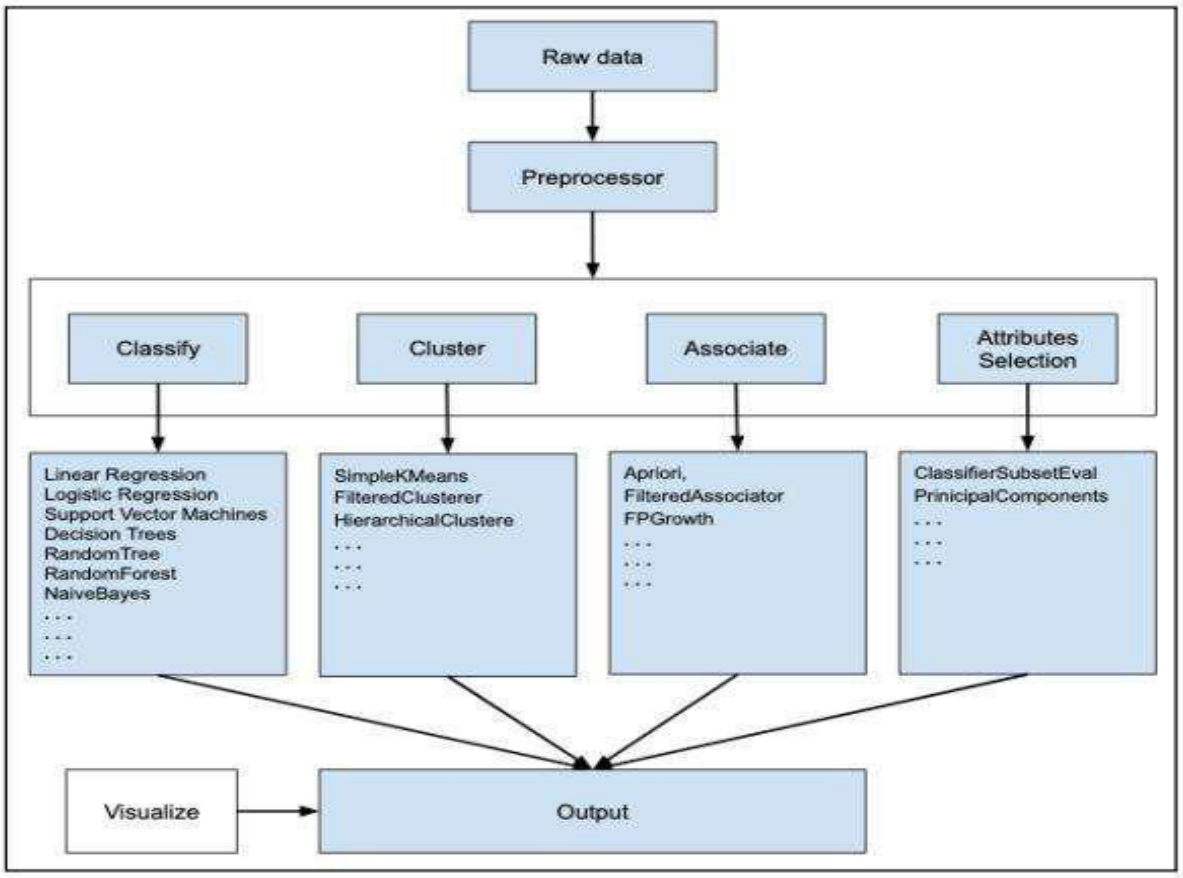
Objective: To make students well versed with open source tool like WEKA to implement Naïve Bayes Classifier.

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:

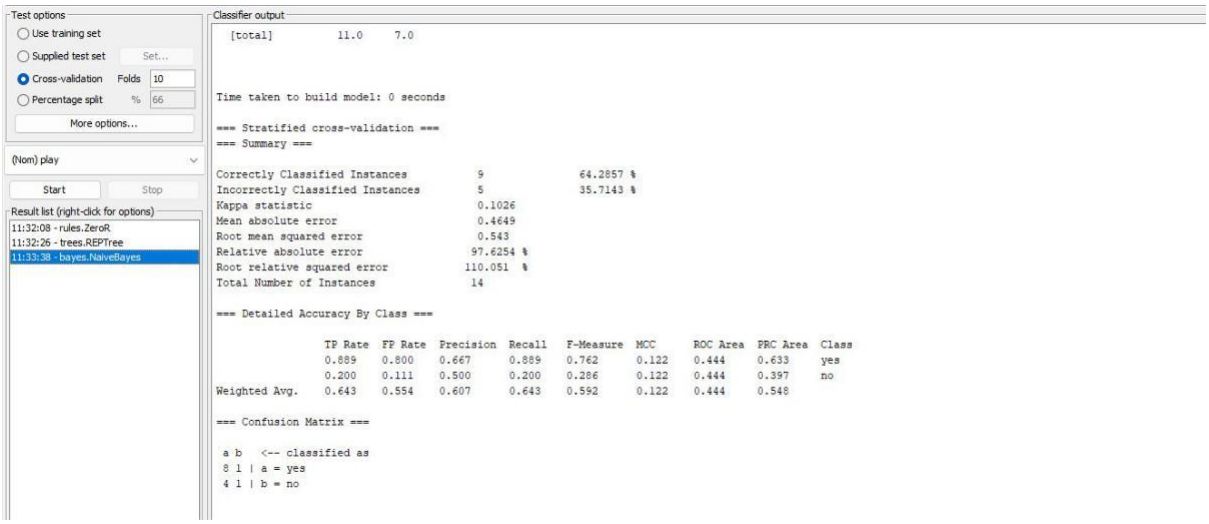
WEKA – an open-source software provides tools for data preprocessing, implementation of several data Mining algorithms, and visualization tools so that you can develop data mining techniques and apply them to real-world data mining problems. Weka is summarized in the following diagram:



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:



Conclusion:

What performance metrics were used to evaluate the Naïve Bayes classifier in WEKA?

The Naïve Bayes classifier in WEKA was evaluated using several key performance metrics. The classifier correctly classified 64.29% of instances, while 35.71% were incorrectly classified. The Kappa statistic, which measures the agreement between predicted and actual classifications beyond chance, was 0.1026. Additionally, the mean absolute error, a measure of the average magnitude of errors in predictions, was 0.4649, and the root mean squared error was 0.543, indicating the overall prediction error. The relative absolute error was 97.63%, while the root relative squared error was 110.05%. These error metrics provide insights into how well the model generalizes across the dataset. A breakdown of detailed accuracy by class shows that for the "yes" class, the true positive (TP) rate was 0.889, and for the "no" class, it was 0.200. Precision, which measures the proportion of correct positive predictions, was
0.667 for "yes" and 0.500 for "no." Recall, the model's ability to correctly identify actual positives, was 0.889 for "yes" and 0.200 for "no." The weighted averages of precision, recall, and other metrics across both classes were also provided. These metrics offer a comprehensive view of the model's classification performance.