Name:	Pratik Sanjay Avhad
Roll No:	01
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.

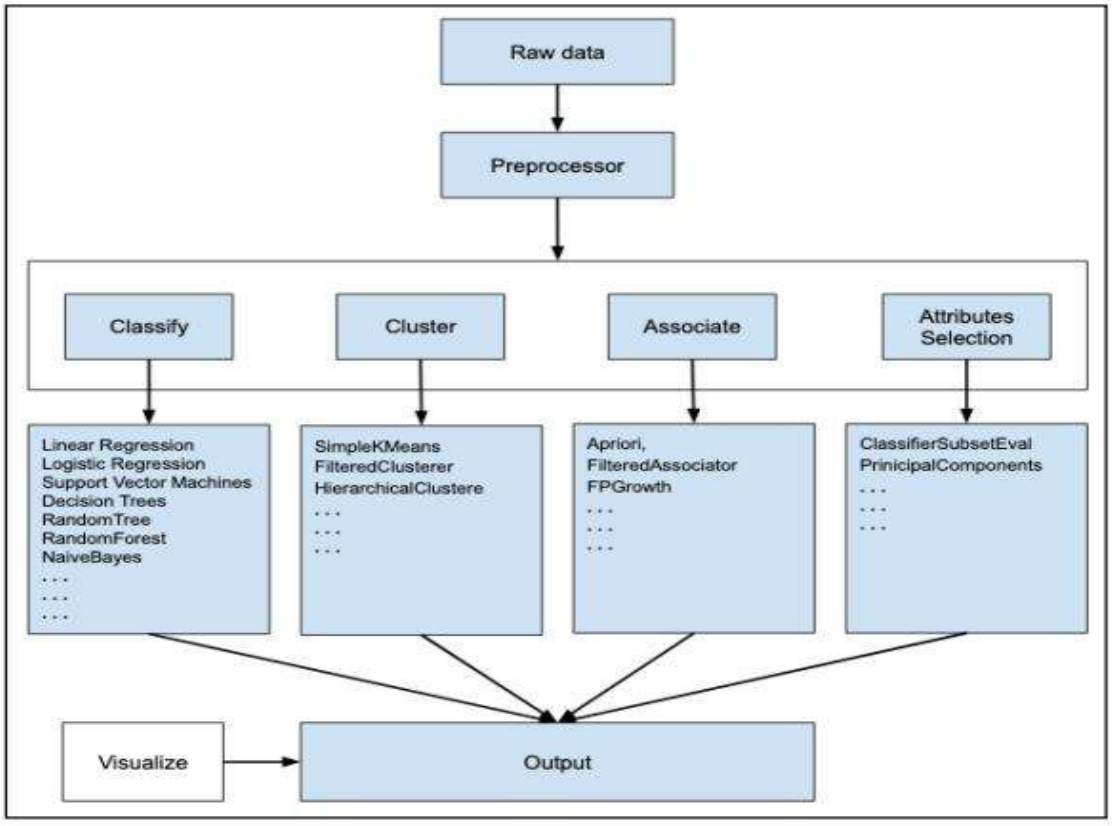
Objective: To make students well versed with open source tool like WEKA to implement Naïve Bayes Classifier.

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:

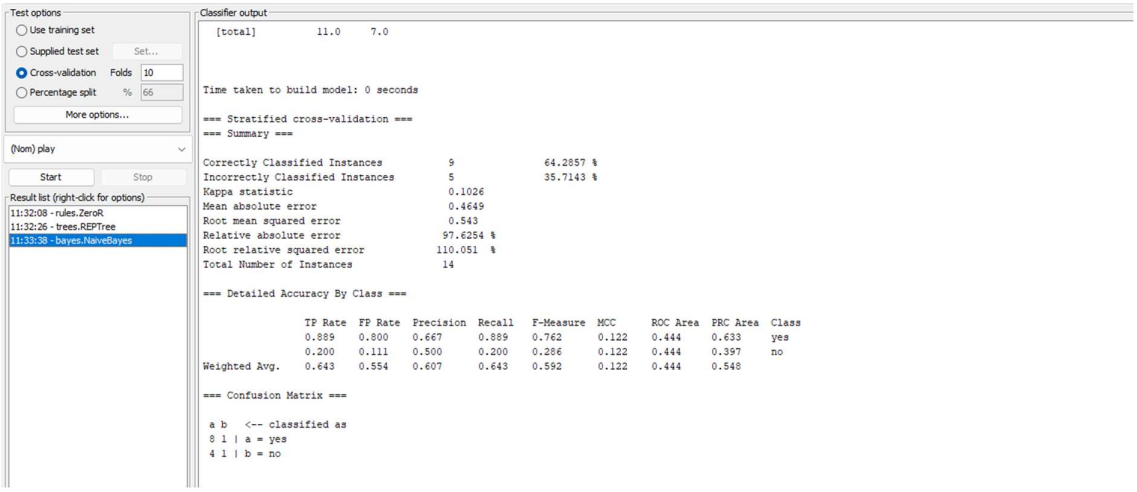
WEKA – an open-source software provides tools for data preprocessing, implementation of several data Mining algorithms, and visualization tools so that you can develop data mining techniques and apply them to real-world data mining problems. Weka is summarized in the following diagram:



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?

In WEKA, several performance metrics are used to evaluate the Naïve Bayes classifier, providing a comprehensive assessment of its effectiveness. **Accuracy** measures the ratio of correctly predicted instances to the total number, but it can be misleading in imbalanced datasets, making additional metrics crucial. **Precision** indicates the accuracy of positive predictions, calculated as the ratio of true positives to total predicted positives, while **recall** (sensitivity) measures the classifier's ability to identify actual positives by comparing true positives to total actual positives. The **F1 score** balances precision and recall, reflecting both the accuracy of positive predictions and the model's ability to capture relevant instances. The **ROC area (AUC)** assesses the model's ability to distinguish between classes, with higher values indicating better performance. Additionally, the **confusion matrix** summarizes prediction results, showing true positives, true negatives, false positives, and false negatives, which aids in visualizing performance across classes. Metrics like the **Kappa statistic** further measure agreement between predicted and actual classifications, accounting for chance. Together, these metrics enable informed

decisions regarding model selection, tuning, and deployment based on specific application requirements.