Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY: 2025-26

Class:	**TE**	**Semester:**	**V**
Course Code:	**CSC504**	**Course Name:**	**Data Warehousing and Mining**

Name of Student:	**Shravani Sandeep Raut**
Roll No. :	**51**
Experiment No.:	**04**
Title of the Experiment:	**Using open source tools Implement Classifiers**
Date of Performance:	
Date of Submission:	

Evaluation

Performance Indicator	**Max. Marks**	**Marks Obtained**
Performance	5	
Understanding	5	
Journal work and timely submission	10	
Total	20	

Performance Indicator	**Exceed Expectations (EE)**	**Meet Expectations (ME)**	**Meet Expect Below Expectations (BE)**
Performance	4-5	2-3	1
Understanding	4-5	2-3	1
Journal work and timely submission	8-10	5-8	1-4

Checked by

Name of Faculty : Ms. Neha Raut

Signature :

Date:

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

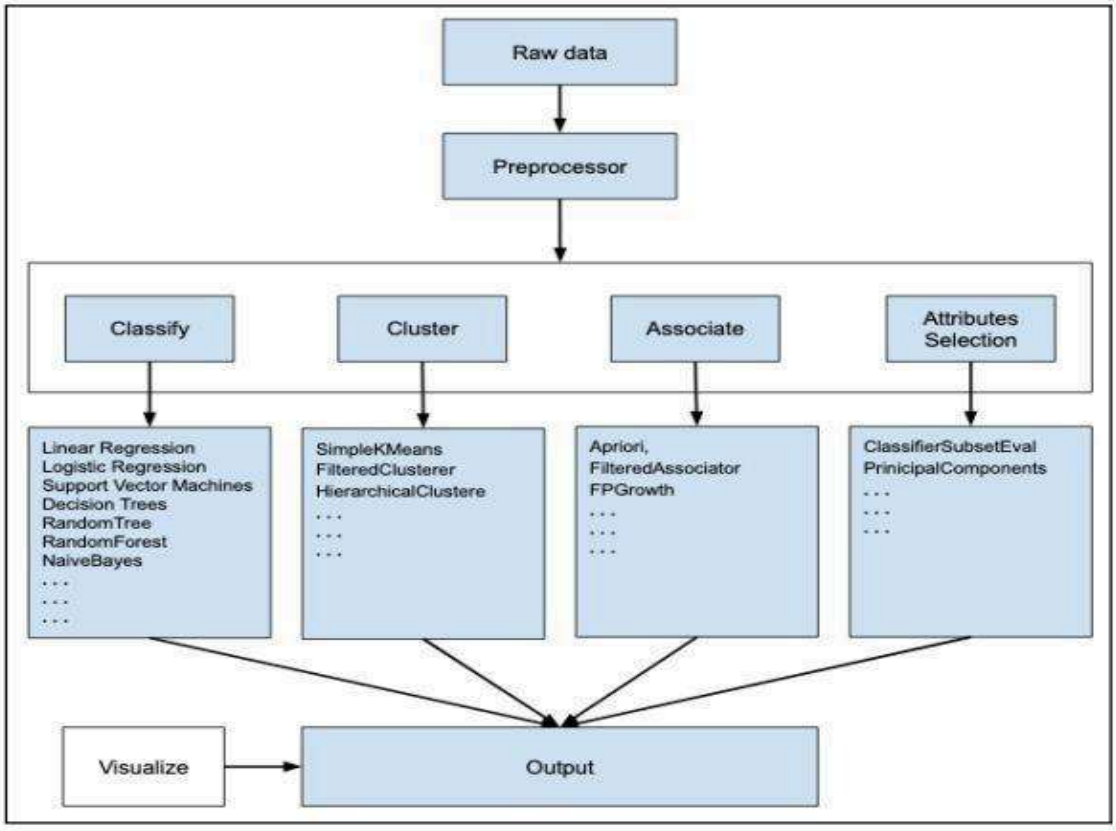
Objective: To make students well versed with open source tool like WEKA to implement Naïve Bayes Classifier.

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:

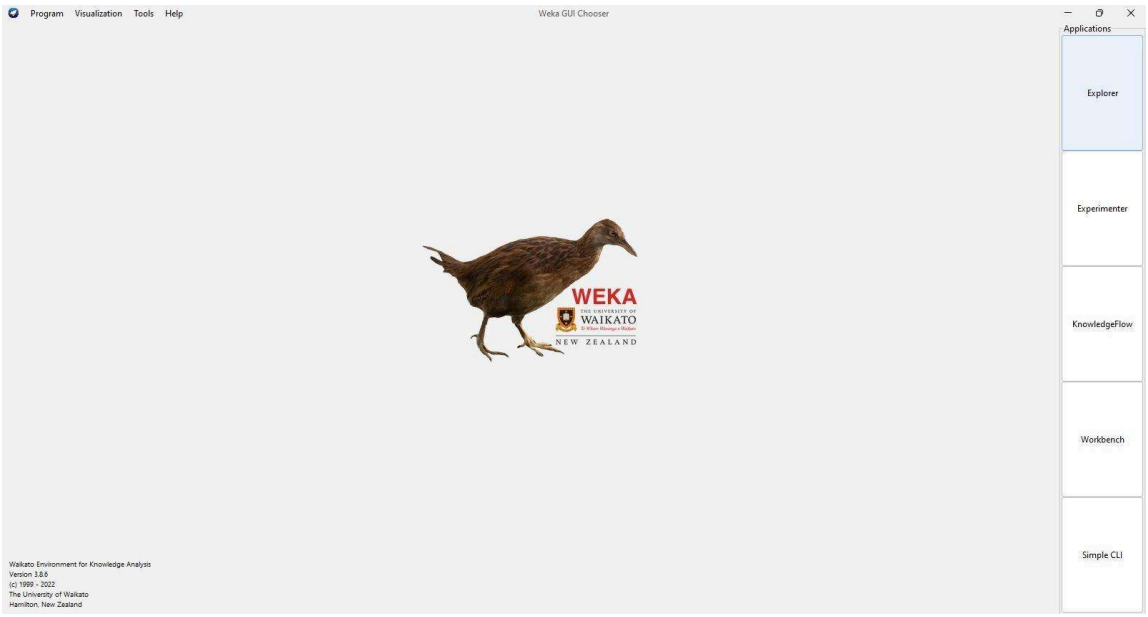


First, you will start with the raw data collected from the field. This data may contain several null values and irrelevant fields. You use the data preprocessing tools provided in WEKA to cleanse the data. Then, you would save the preprocessed data in your local storage for applying Data Mining algorithms.

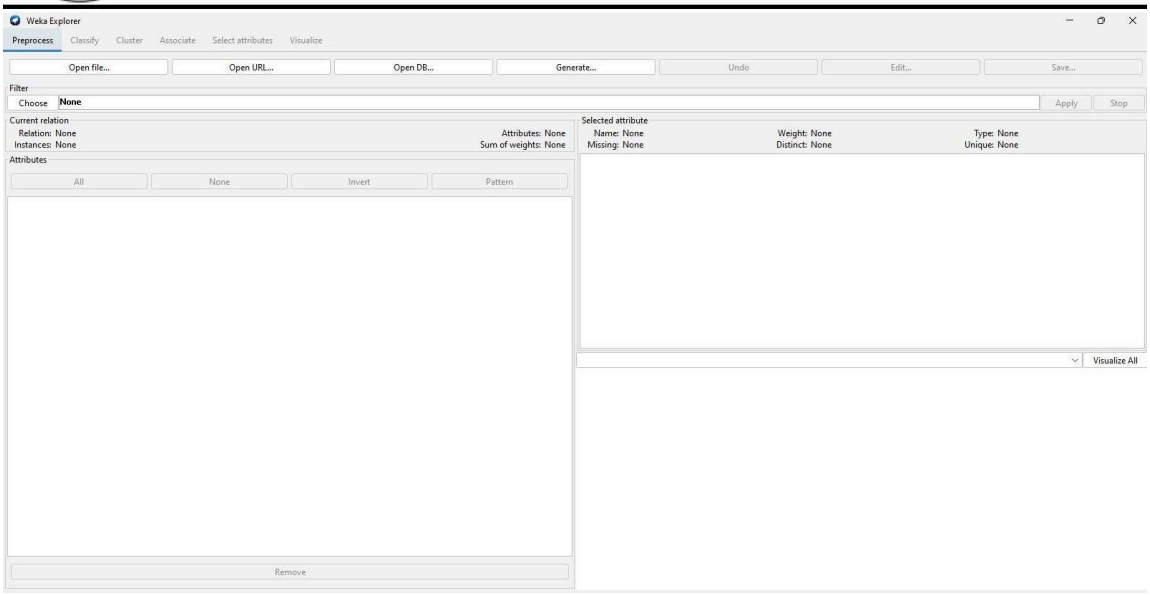
Next, depending on the kind of Data Mining model that you are trying to develop you would select one of the options such as Classify, Cluster, or Associate. The Attributes Selection allows the automatic selection of features to create a reduced dataset. Note that under each category, WEKA provides the 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:

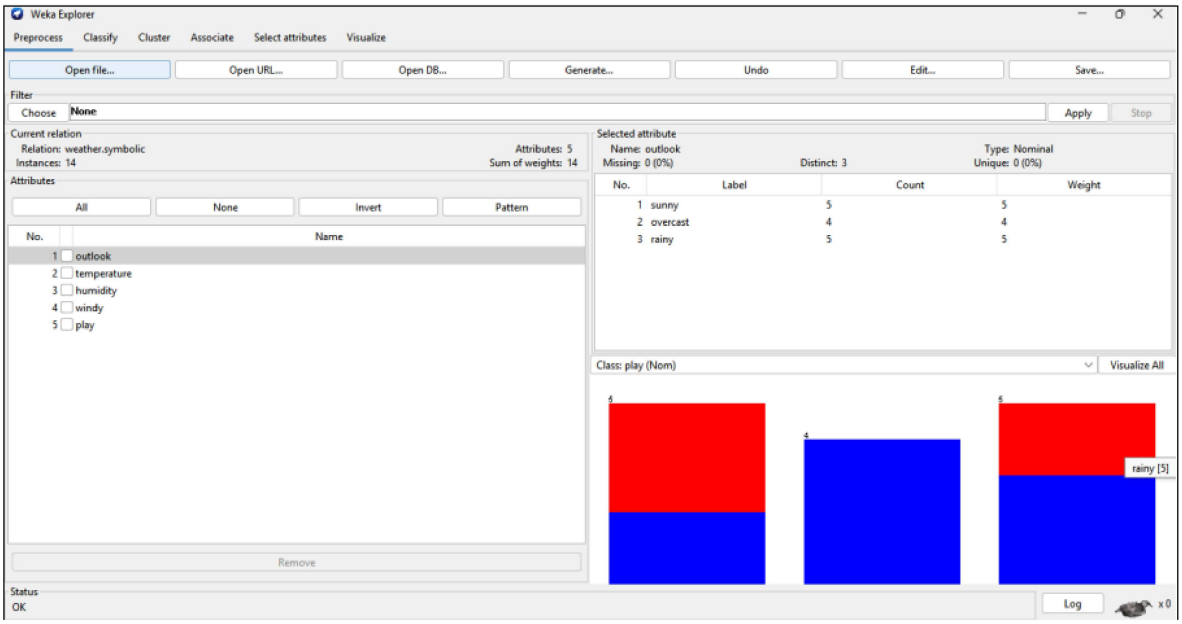
Step 1:Open WEKA, the following GUI should appear on your system.



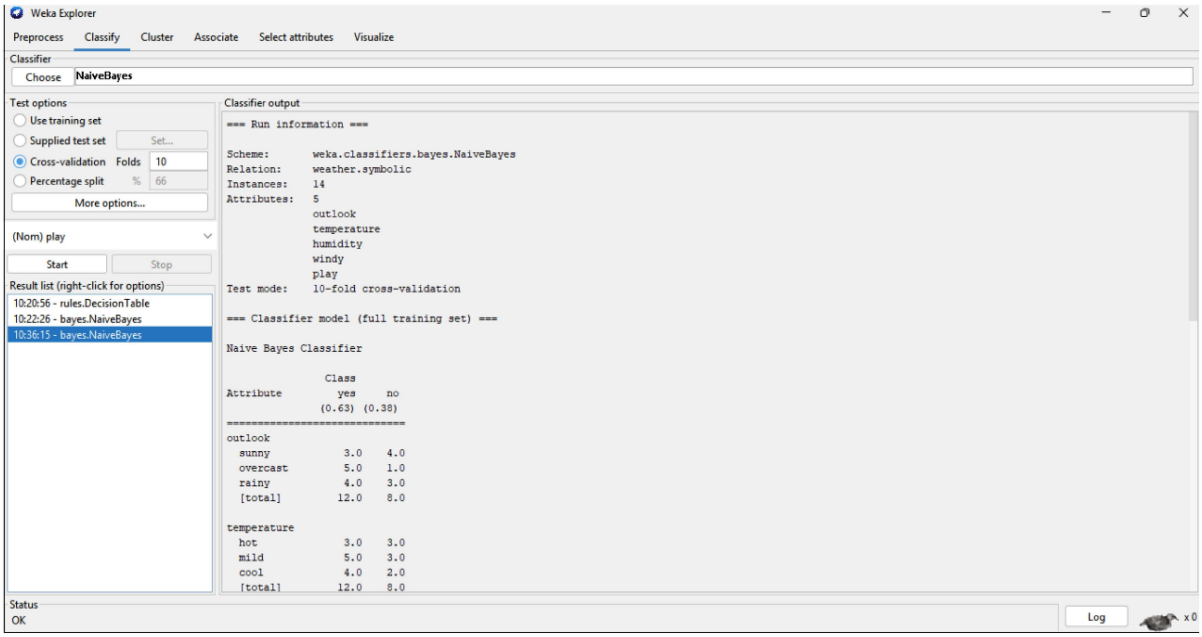
Step 2: Click on the explorer.



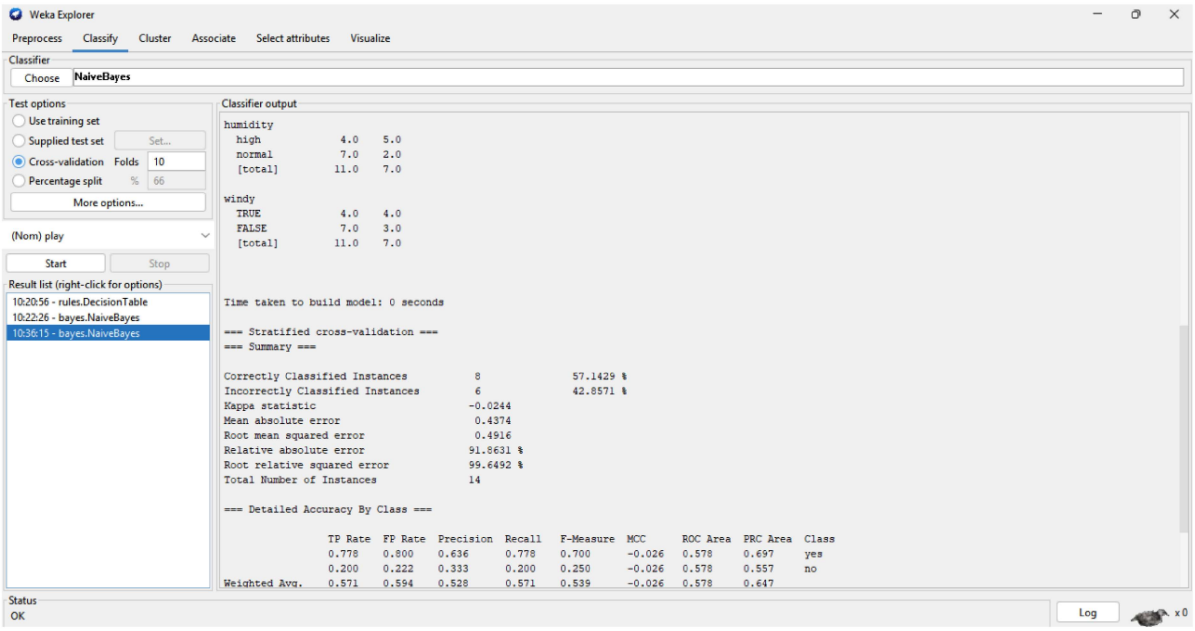
Step 3: Select 'weather.numeric.arff' dataset which already exists in the program files, from the 'Open File' section. The following screen should appear.



Step 4: Select 'Naive bayes classifier', under classifiers->Bayes->Naive Bayes and click on 'Start'.



Step 5: Click on 'Start' , the following output will appear.



Conclusion:

What performance metrics were used to evaluate the Naïve Bayes classifier in WEKA?In WEKA, the performance metrics commonly used to evaluate the Naïve Bayes classifier (or any classifier) typically include:

1. **Accuracy**: The proportion of correctly classified instances out of the total instances.
2. **Precision**: The ratio of true positive predictions to the total predicted positives, indicating the quality of the positive class predictions.
3. **Recall (Sensitivity)**: The ratio of true positive predictions to the actual positives, measuring the ability of the classifier to find all positive instances.
4. **F1 Score**: The harmonic mean of precision and recall, providing a balance between the two metrics.
5. **ROC-AUC (Receiver Operating Characteristic - Area Under Curve)**: A performance measurement for classification problems at various threshold settings, assessing the classifier's ability to distinguish between classes.
6. **Confusion Matrix**: A table that summarizes the performance of the classifier, showing true positives, true negatives, false positives, and false negatives.
7. **Kappa Statistic**: A measure of how much better the classifier performs compared to random guessing, taking into account the agreement occurring by chance.

These metrics can be accessed through the evaluation results in WEKA after running the aïve Bayes classifier on a dataset