Academic Year: 2025-26 **Semester: V**
Class/Branch: T.E. DS **Subject: DWMLab**

EXPERIMENT NO. 7

1. **Aim:** To study the WEKA environment and analyze the effectiveness of the Decision Tree and Naive Bayes algorithms for classification.

2. **Objectives:** From this experiment, the student will be able to

   - Learn about the Weka Data Mining tool.
   - To implement Decision Tree classifier-J48 using Weka.
   - To implement the Naïve Bayes classifier using Weka.

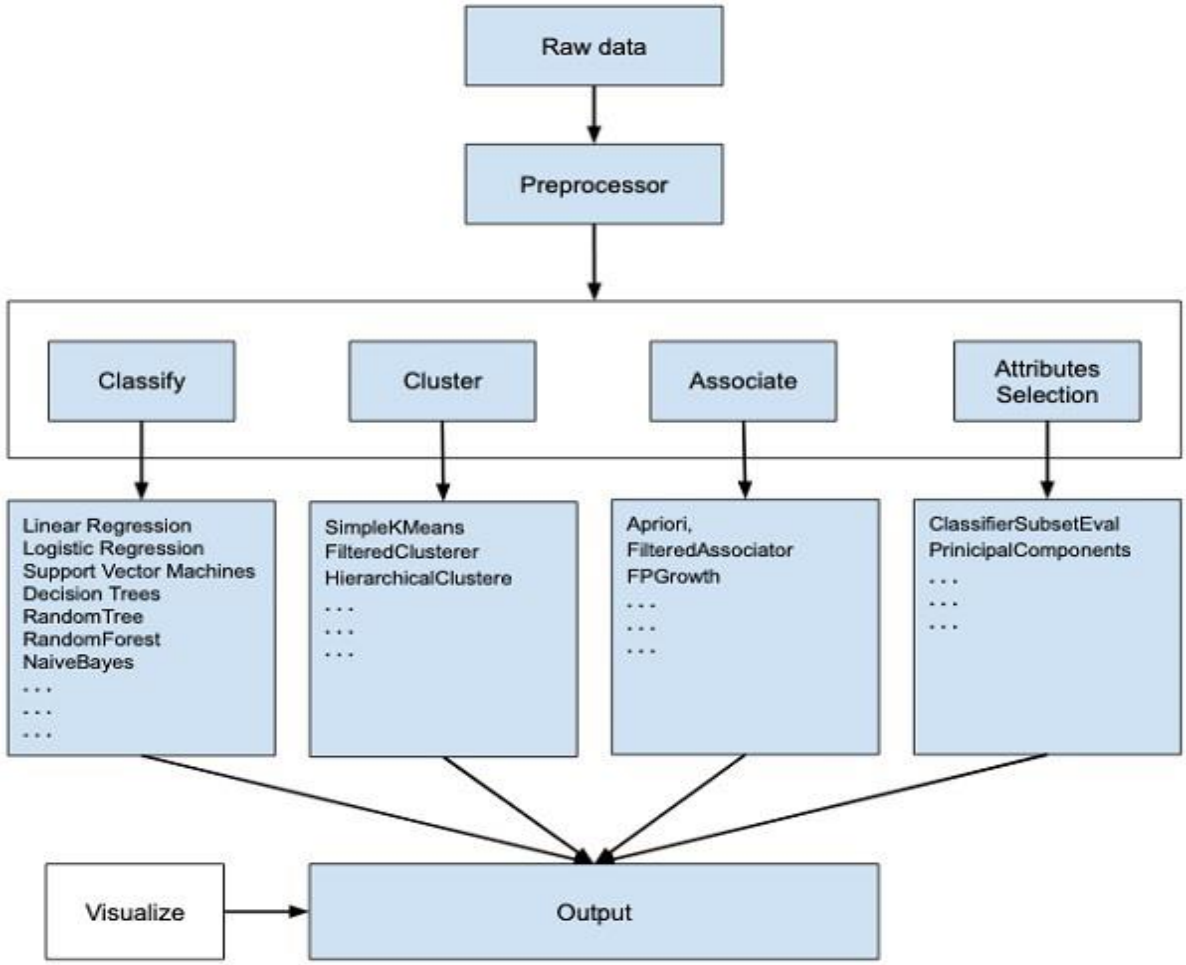
3. **Theory:**

   Waikato Environment for Knowledge Analysis - WEKA - an open source software provides tools for data pre-processing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real- world data mining problems.
   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 pre-processing tools provided in WEKA to cleanse the data.

   Next, depending on the kind of ML 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.

   What WEKA offers is summarized in the following diagram –



Pre-processing using WEKA:

Using the **Open file** ... option under the **Pre-process** tag select the **labor.arff** file.

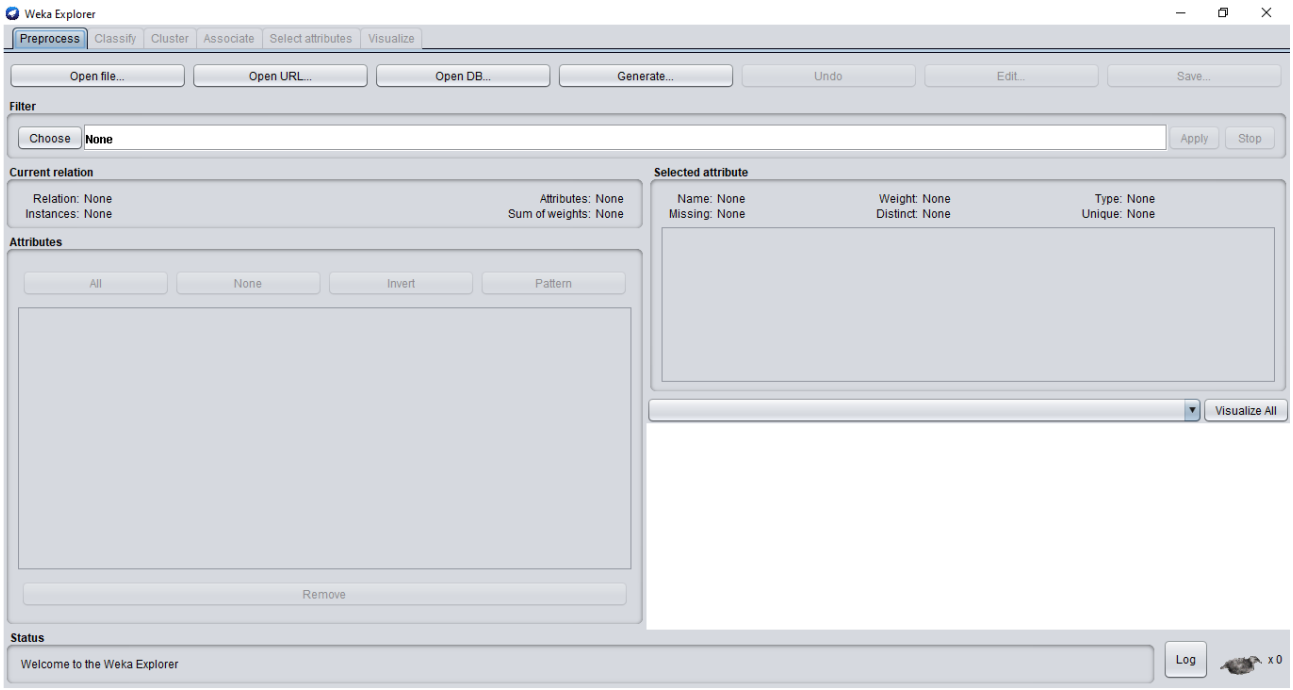
Using Filters:

weka→filters→unsupervised→attribute→ReplaceWithMissingValues

Classification using WEKA:

Decision tree:

A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision. Decision tree are commonly used for gaining information for the purpose of decision -making.

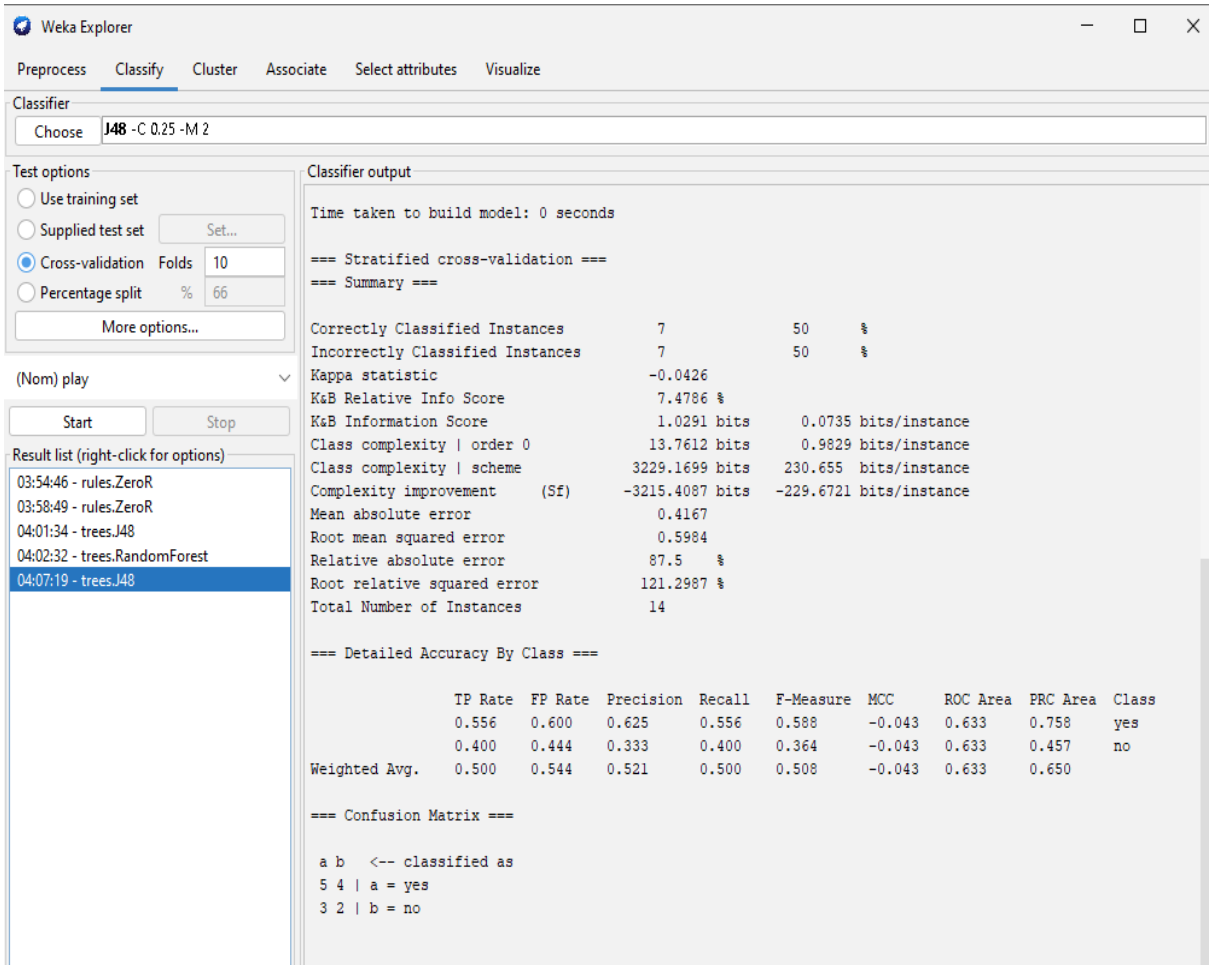


This experiment illustrates the use of j-48 classifier in weka. The sample data set used in this experiment is "weather.arff" data available arff format.

Selecting Classifier :j-48 classifier in weka

Click on the Choose button and select the following classifier –

weka→classifiers>trees>J48



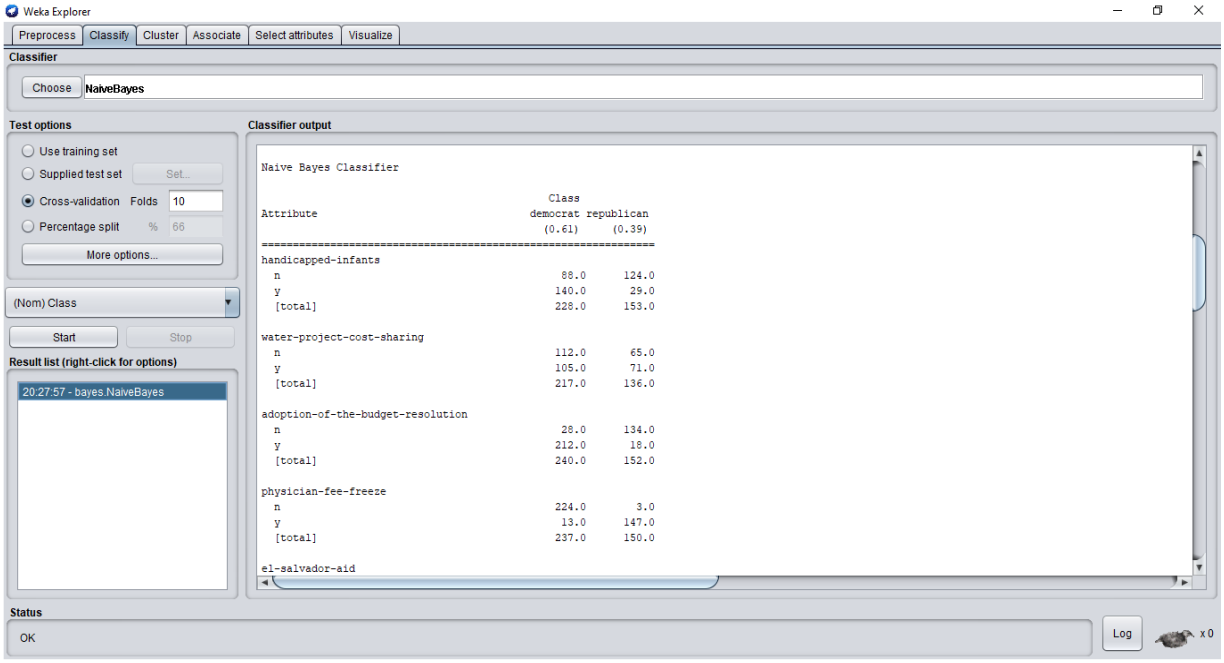
Naive Bayes:

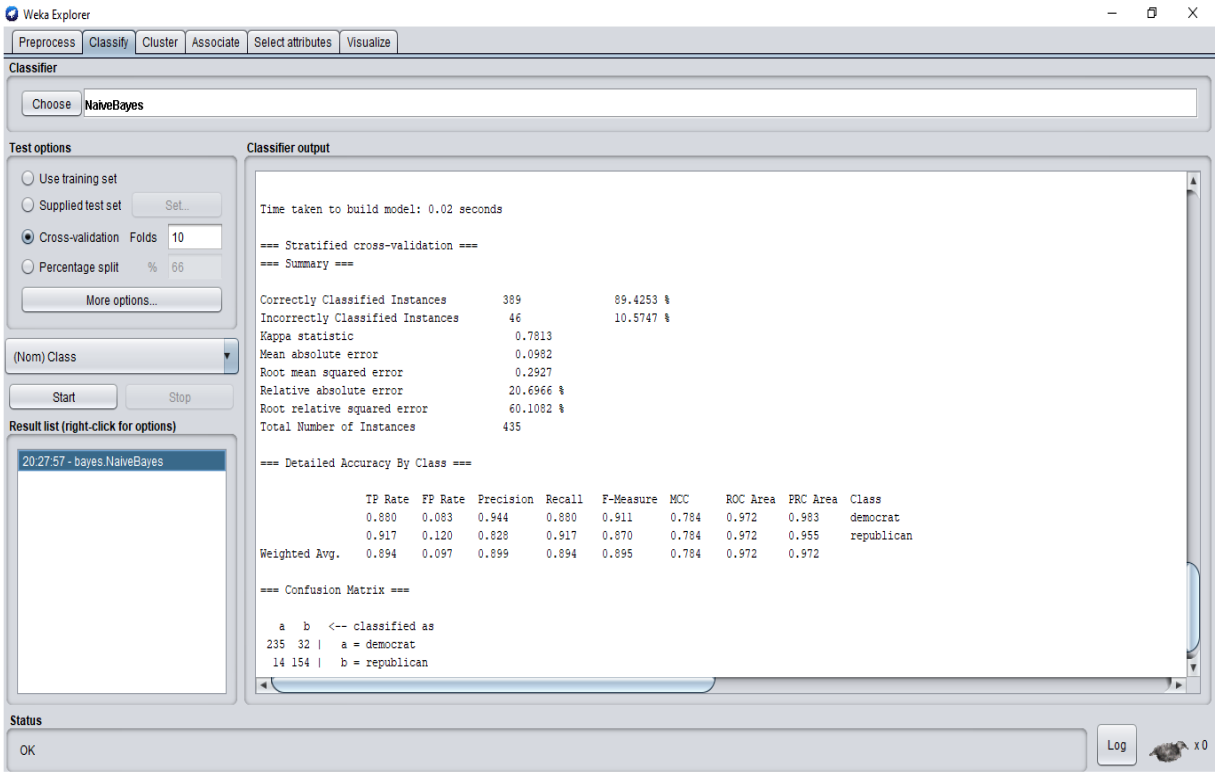
Naive Bayes uses a simple implementation of Bayes Theorem (hence naive) where the prior probability for each class is calculated from the training data and assumed to be independent of each other (technically called conditionally independent).

Selecting Classifier

Click on the Choose button and select the following classifier –

weka→classifiers>bayes>Naïve Bayes





4. Conclusion: