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

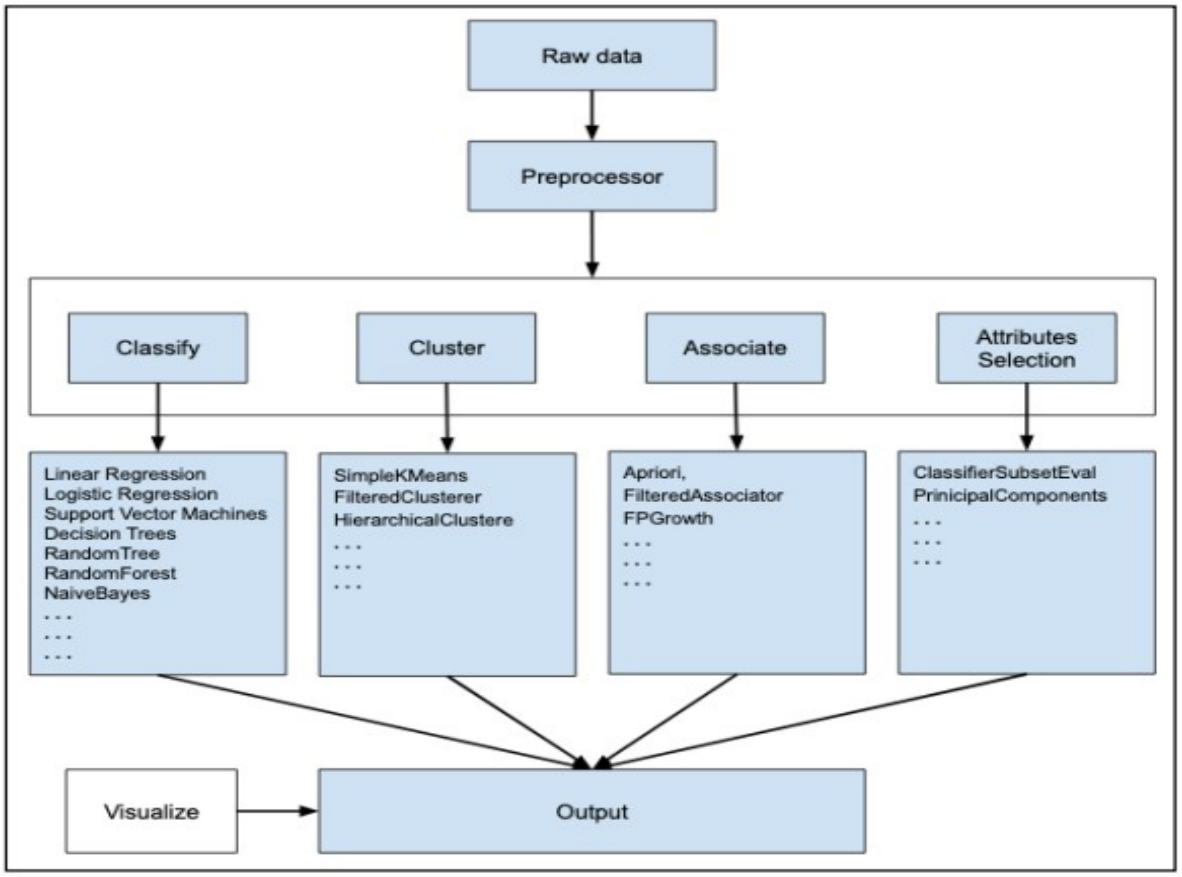
**Objective:** To make students well versed with open source tool like WEKA to implement Naïve Bayes Classifier.

**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:**

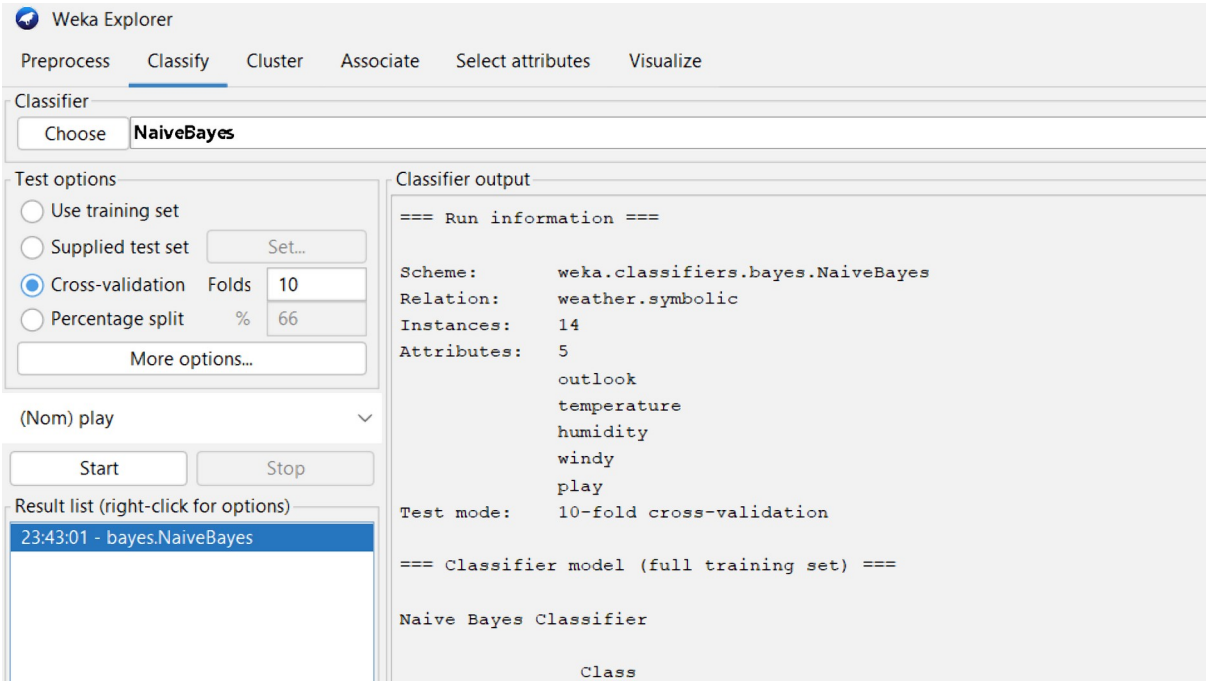
WEKA – an open-source software provides tools for data preprocessing, implementation of several data Mining algorithms, and visualization tools so that you can develop data mining techniques and apply them to real-world data mining problems. Weka is summarized in the following diagram:




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:**



```
Naive Bayes Classifier

                Class
Attribute         yes     no
                (0.63) (0.38)
============================
outlook
  sunny            3.0    4.0
  overcast         5.0    1.0
  rainy            4.0    3.0
  [total]         12.0    8.0

temperature
  hot              3.0    3.0
  mild             5.0    3.0
  cool             4.0    2.0
  [total]         12.0    8.0

humidity
  high             4.0    5.0
  normal           7.0    2.0
  [total]         11.0    7.0
```

Status
OK

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier
Choose NaiveBayes

Test options
- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) play

Start | Stop

Result list (right-click for options)
23:43:01 - bayes.NaiveBayes

Classifier output

```
  normal           7.0    2.0
  [total]         11.0    7.0

windy
  TRUE             4.0    4.0
  FALSE            7.0    3.0
  [total]         11.0    7.0



Time taken to build model: 0 seconds

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

Correctly Classified Instances           8               57.1429 %
Incorrectly Classified Instances         6               42.8571 %
Kappa statistic                         -0.0244
```

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

Correctly Classified Instances           8               57.1429 %
Incorrectly Classified Instances         6               42.8571 %
Kappa statistic                         -0.0244
Mean absolute error                      0.4374
Root mean squared error                  0.4916
Relative absolute error                 91.8631 %
Root relative squared error             99.6492 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.778    0.800    0.636      0.778    0.700      -0.026   0.578     0.697     yes
                 0.200    0.222    0.333      0.200    0.250      -0.026   0.578     0.557     no
Weighted Avg.    0.571    0.594    0.528      0.571    0.539      -0.026   0.578     0.647

=== Confusion Matrix ===

 a b   <-- classified as
 7 2 | a = yes
 4 1 | b = no
```

**Conclusion:**
In this practical, we successfully implemented the Naïve Bayes classifier using the open-source tool WEKA. We explored how the algorithm works by classifying data and interpreting the results, such as accuracy and the confusion matrix. This helped us to understand how to use WEKA for data mining tasks and how Naïve Bayes can be used for classification problems.

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

**Correctly Classified Instances**: This metric shows the percentage of instances that the model correctly predicted. In your case, 57.14% of the instances were correctly classified.

**Incorrectly Classified Instances**: This shows the percentage of instances that the model classified incorrectly, which is 42.86%.

**Kappa Statistic**: This statistic measures how much better the classifier performed compared to a random classifier. A value of -0.0244 indicates that the model performed slightly worse than random chance.

**Mean Absolute Error (MAE)**: The average prediction error is 0.44, meaning on average, the model's predictions are off by 0.44.

**Root Mean Squared Error (RMSE)**: The model's larger errors are slightly more penalized, with an average error of about 0.49.

**Relative Absolute Error (RAE)**: The model's error is 91.86% of what you'd expect from a simple baseline, meaning it's only slightly better than guessing the average.

**Root Relative Squared Error (RRSE)**: The model's squared error is almost the same as a baseline predictor, at 99.65%.

**Total Number of Instances**: The model was tested on 14 data points.