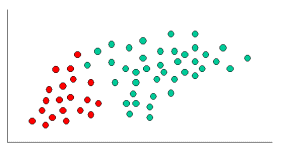
**Experiment No**

**AIM:**To study Classification Algorithm in weka.

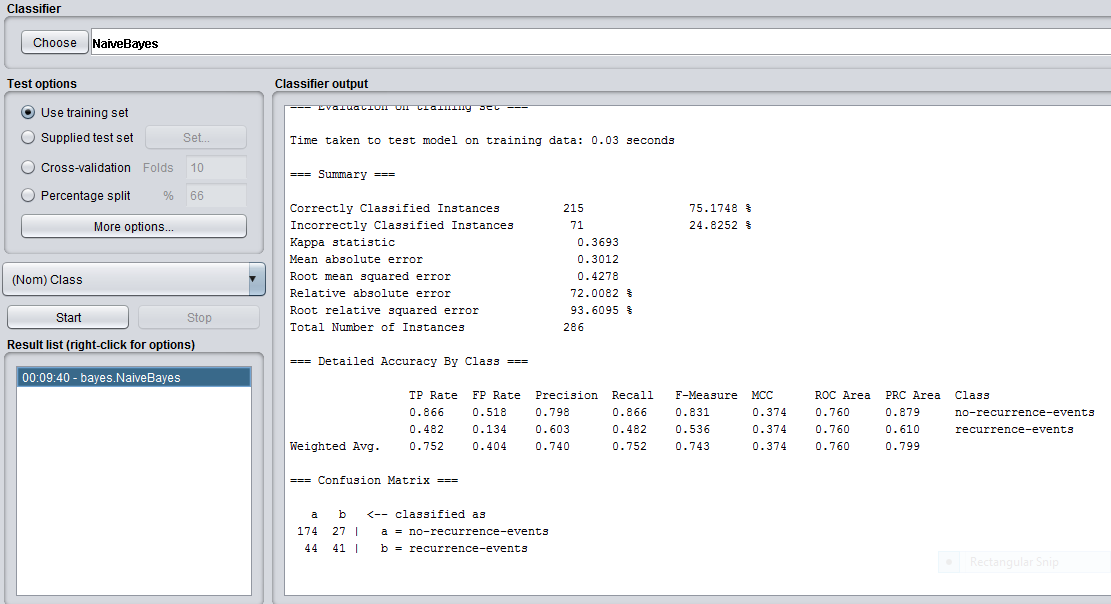
Theory:

The Naive Bayes Classifier technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods.

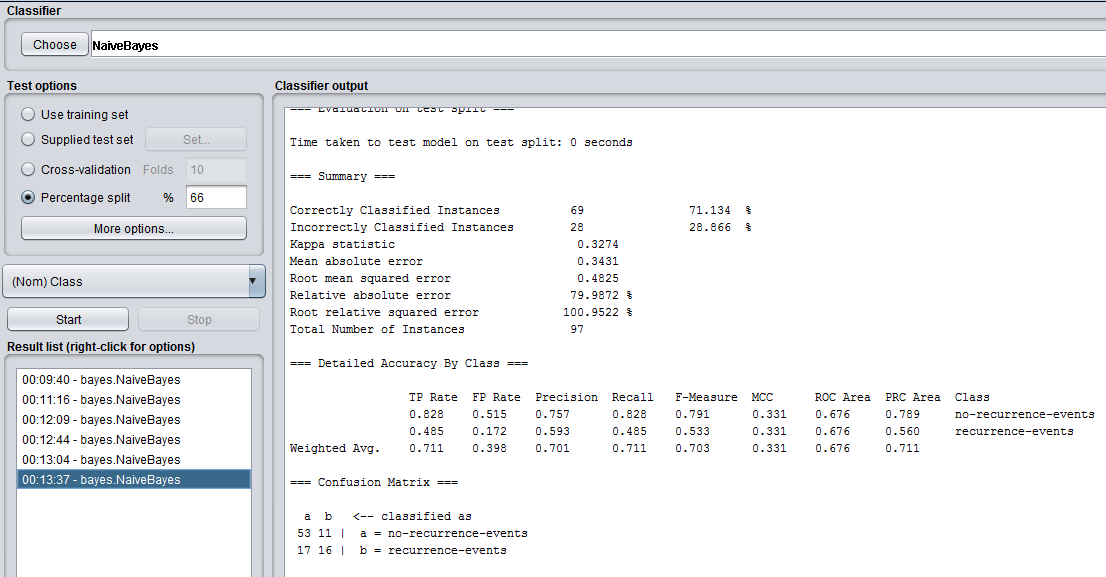


To demonstrate the concept of Naïve Bayes Classification, consider the example displayed in the illustration above. As indicated, the objects can be classified as either GREEN or RED. Our task is to classify new cases as they arrive, i.e., decide to which class label they belong, based on the currently existing objects.

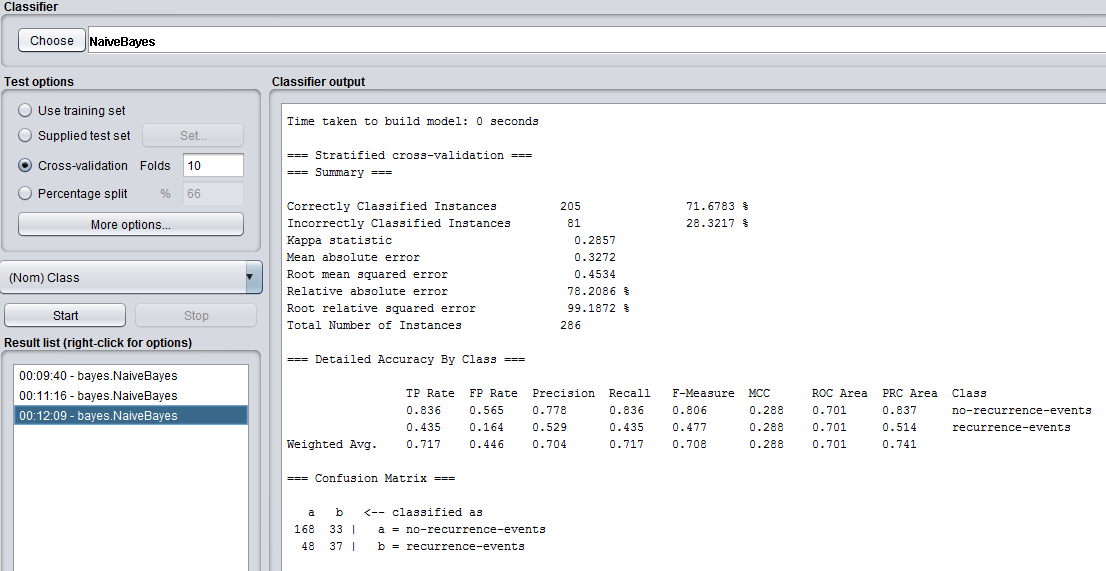
Since there are twice as many GREEN objects as RED, it is reasonable to believe that a new case (which hasn't been observed yet) is twice as likely to have membership GREEN rather than RED. In the Bayesian analysis, this belief is known as the prior probability. Prior probabilities are based on previous experience, in this case the percentage of GREEN and RED objects, and often used to predict outcomes before they actually happen.

**1] Using training set (default).** After generating the classification Weka classifies the training instances into classification according to the label representation and computes the percentage of instances falling in each classification.

**2] Using Percentage Split:**If you have a data of 286 instances and you would like to split 66% as training and 34% as test set using percentage split (order preserved). The 66th instance will be as training and the 67th until the last instance will be taken as test set.



3)cross-validation:**Cross**-**validation** is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. In (10)k-fold **cross**-**validation**, the original sample is randomly partitioned into k equal size subsamples.



Conclusion:on given dataset training set technique for testing perform better than cross validation and percentage split with accuracy of 75% and parameter ID,Name,class are best for finding classification.