



Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

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.

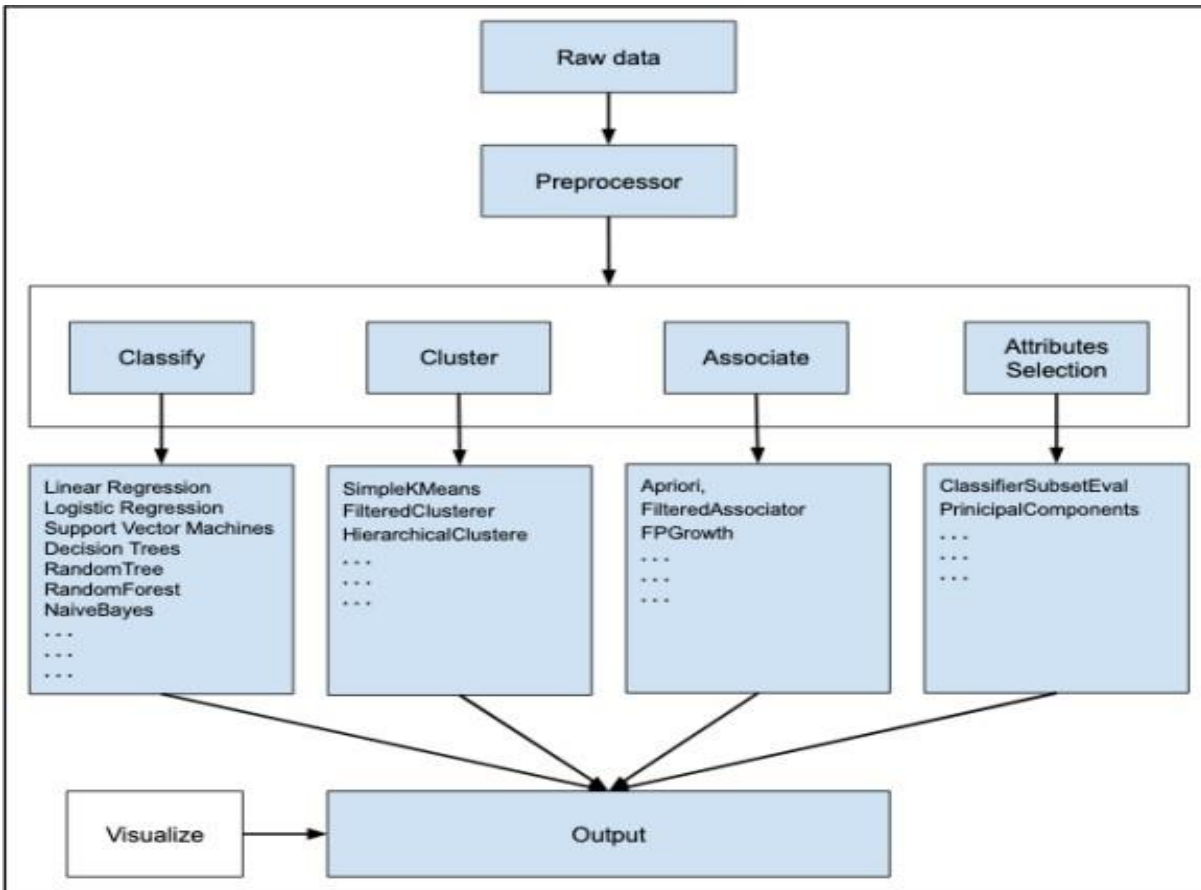
WEKA:

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:



Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science



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.

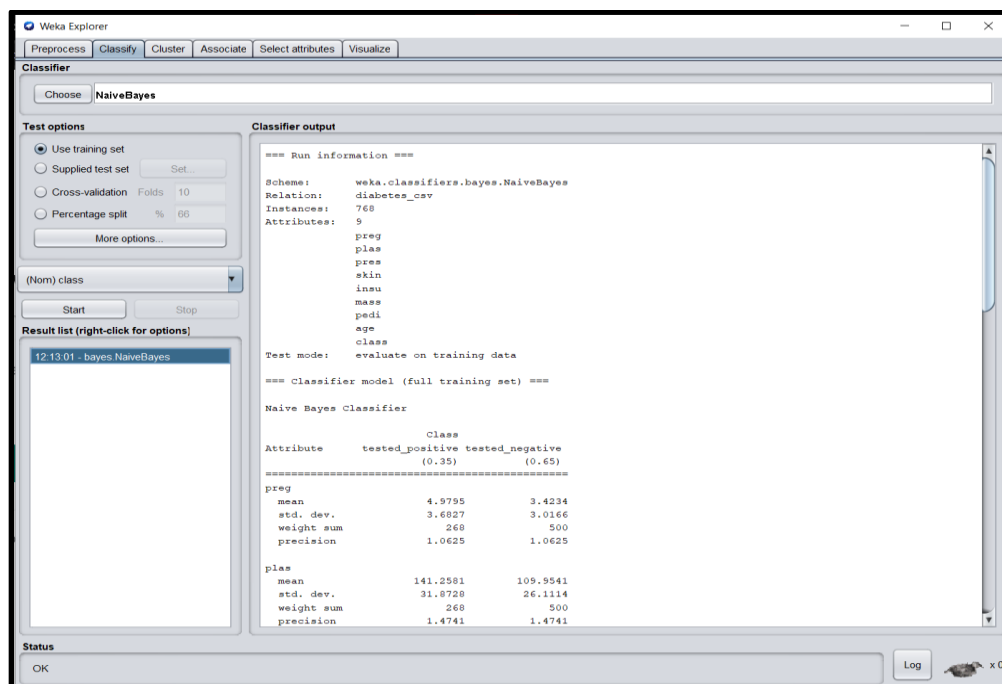
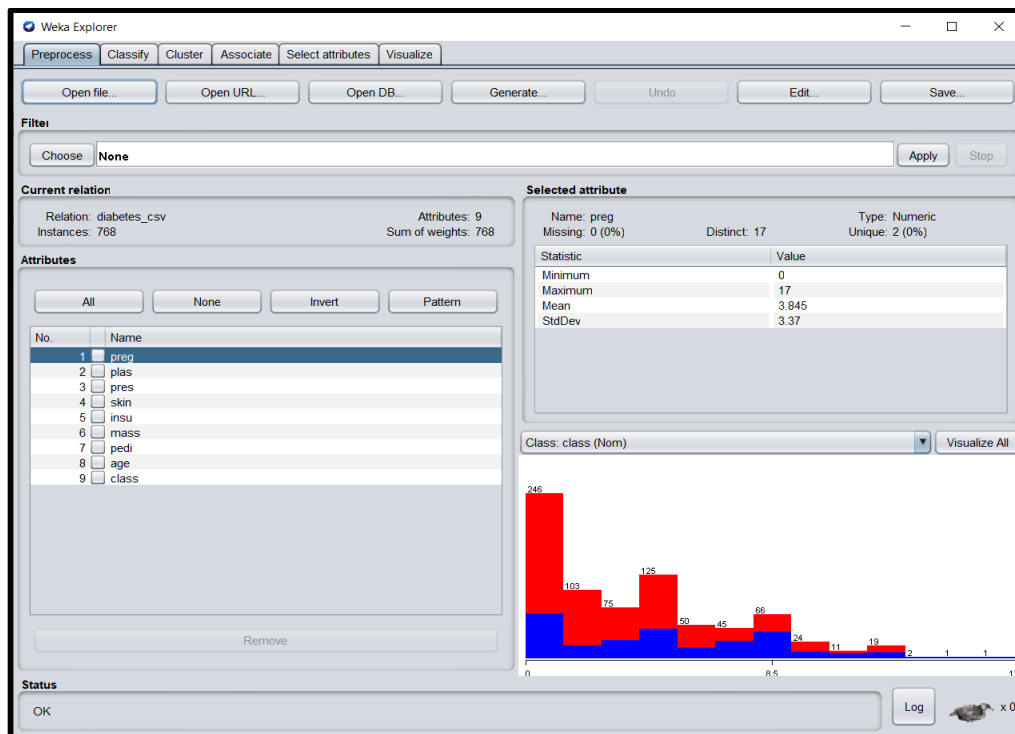


Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

Output:

WEKA TOOL





Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

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) class

Start Stop

Result list (right-click for options)

12.13.01 - bayes NaiveBayes

Classifier output

pres			
mean	70.718	68.1397	
std. dev.	21.4094	17.9834	
weight sum	268	500	
precision	2.6522	2.6522	
skin			
mean	22.2824	19.8356	
std. dev.	17.6992	14.8974	
weight sum	268	500	
precision	1.98	1.98	
insu			
mean	100.2812	68.8507	
std. dev.	138.4883	98.828	
weight sum	268	500	
precision	4.573	4.573	
mass			
mean	35.1475	30.3009	
std. dev.	7.2537	7.6833	
weight sum	268	500	
precision	0.2717	0.2717	
pedi			
mean	0.5504	0.4297	
std. dev.	0.3715	0.2966	
weight sum	268	500	
precision	0.0045	0.0045	
age			
mean	37.0808	31.2494	
std. dev.	10.9146	11.6059	
weight sum	268	500	
precision	1.1765	1.1765	

Status

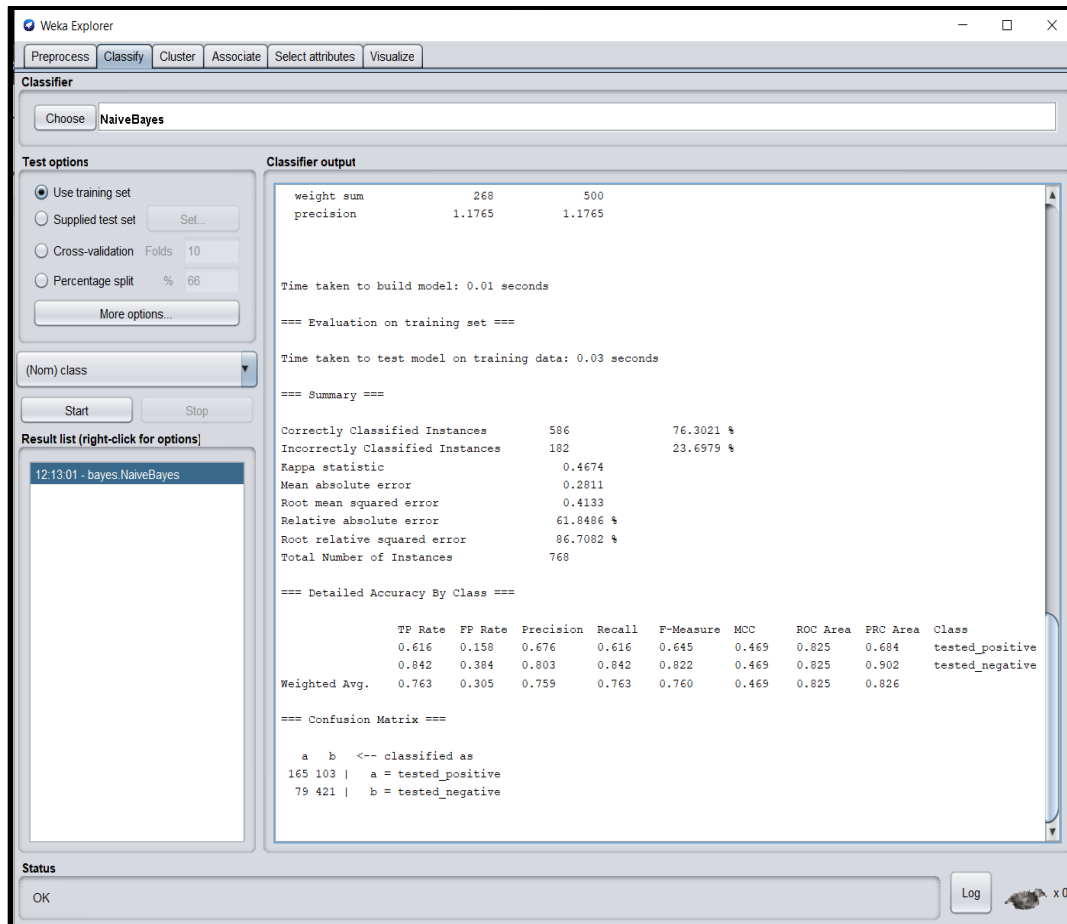
OK Log x 0



Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

WEKA OUTPUT



Naive Bayes model accuracy (in %):

76.62337662337663 (Using Python)
76.3021 (Using Weka)

Conclusion:

We observed that Bayesian Classification represents a supervised learning method as well as a statistical classification method. The Bayesian classification is used as a probabilistic learning method as Naive Bayes classifiers are among the most successful known algorithms for learning to classify the datasets. Naïve Bayes classification Algorithm is one of the probabilistic algorithms which classifies the datasets according to its knowledge data and creates the Result as per the given knowledge.

- Naïve Bayes classification Algorithm is one of the probabilistic algorithms which classify the datasets according to their knowledge data and creates the Result as per the given knowledge.
- Hence we've successfully implemented the Naive Bayesian Algorithm through Python as well as Weka Tool.