Name:	Prerna Sanjay Mhatre
Roll No:	68
Class/Sem:	TE/V
Experiment No.:	4
Title:	Using open source tools Implement Classifiers
Date of Performance:	
Date of Submission:	
Marks:	
Sign of Faculty:	

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

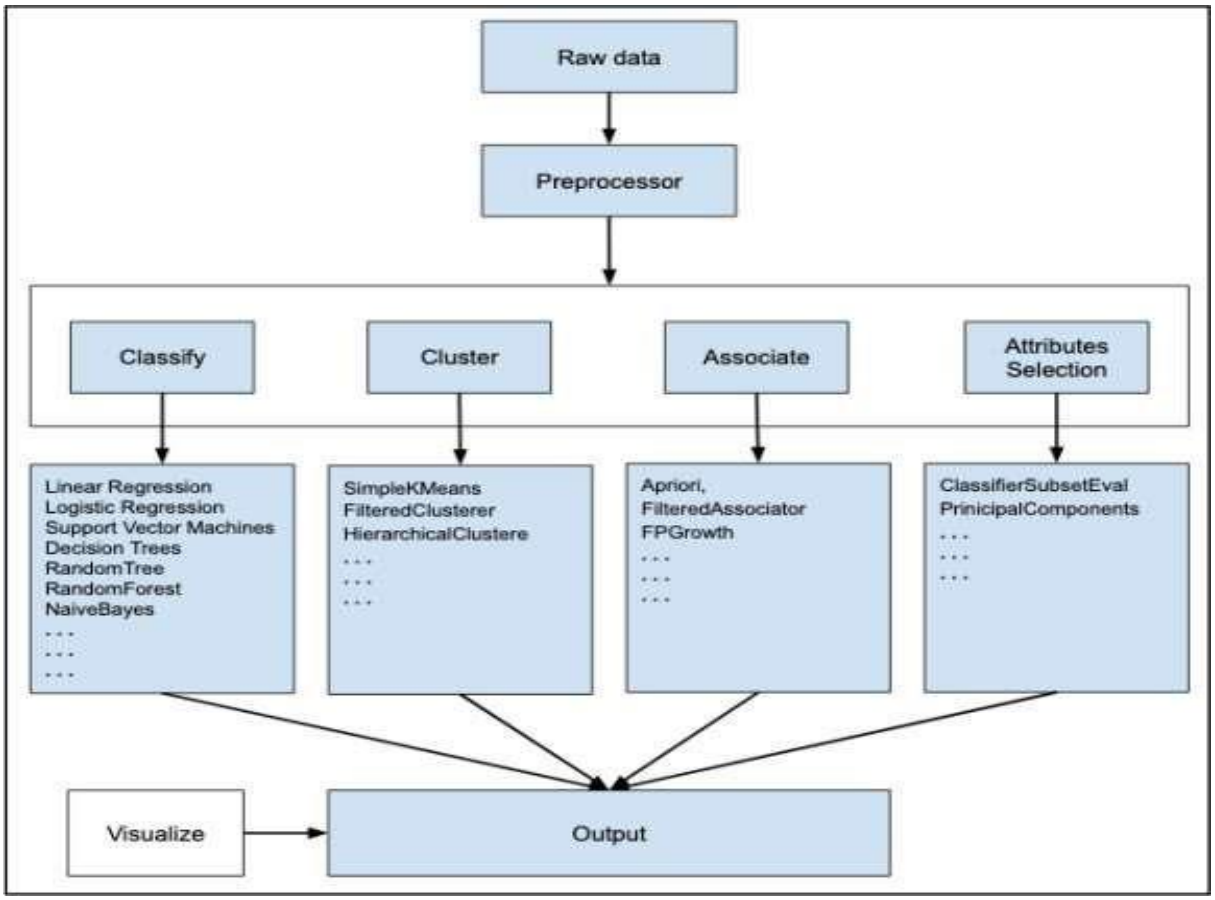
Objective: To make students well versed with open source tool like WEKA to implement Naïve Bayes Classifier.

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:



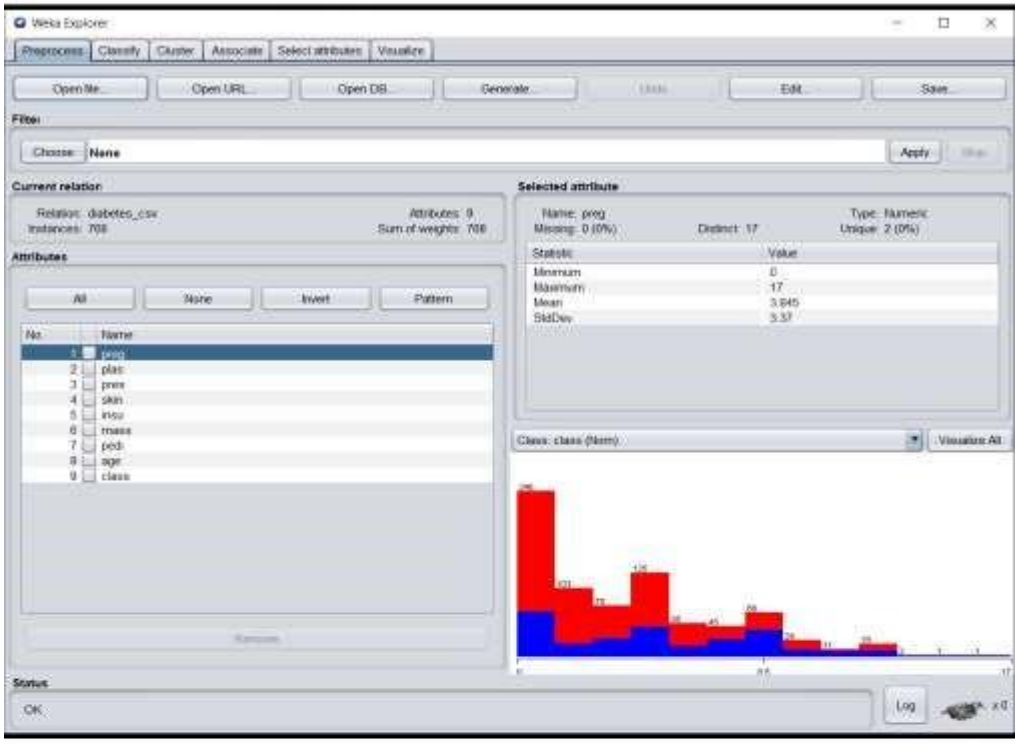
First, you will start with the raw data collected from the field. This data may contain several null values and irrelevant fields. You use the data preprocessing tools provided in WEKA to cleanse the data. Then, you would save the preprocessed data in your local storage for applying Data Mining algorithms.

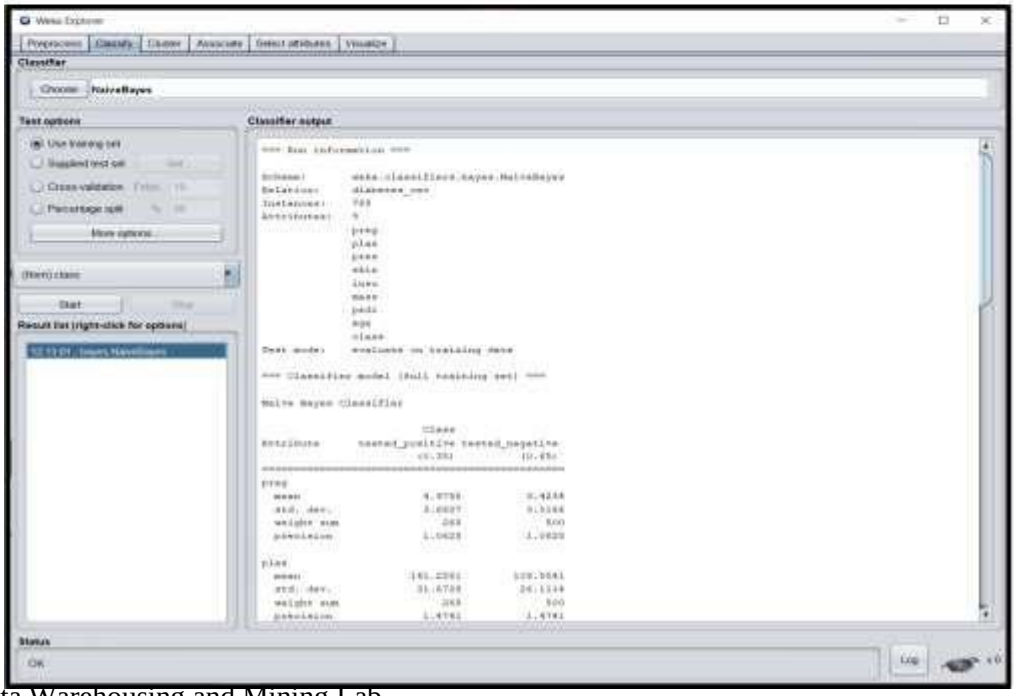
Next, depending on the kind of Data Mining model that you are trying to develop you would select one of the options such as Classify, Cluster, or Associate. The Attributes Selection allows the automatic selection of features to create a reduced dataset. Note that under each category, WEKA provides the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

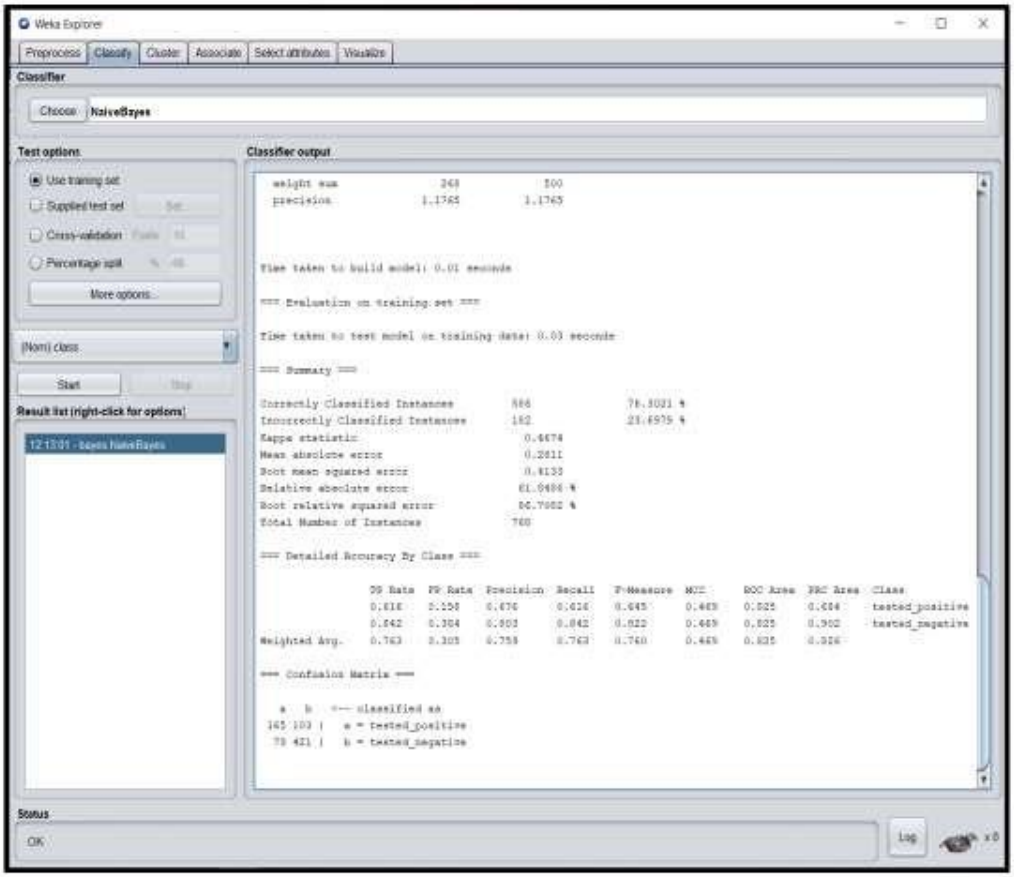
WEKA TOOL :


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: diabetes_csv
Instances: 768
Attributes: 9
Sum of weights: 768
Selected attribute
Name: preg
Missing: 0 (0%)
Distinct: 17
Type: Numeric
Unique: 2 (0%)
Attributes
All
None
Invert
Pattern
No.
Name
preg
plas
pres
skin
insu
mass
pedi
age
class
Statistic
Value
Minimum
0
Maximum
17
Mean
3.845
StdDev
3.37
Class: class (Nom)
Visualize All
Status
OK
Log


Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More options...
Start
Stop
Result list (right-click for options)
Classifier output
=== Run information ===
Scheme: weka.classifiers.bayes.NaiveBayes
Relation: diabetes_csv
Instances: 768
Attributes: 9
preg
plas
pres
skin
insu
mass
pedi
age
class
Test mode: evaluate on training data
=== Classifier model (full training set) ===
Naive Bayes Classifier
Class
Attribute tested_positive tested_negative
(0.35) (0.65)
preg
mean 4.8756 3.4258
std. dev. 3.0627 3.1164
weight sum 268 500
precision 1.0628 1.0628
plas
mean 145.2351 109.9541
std. dev. 31.6726 26.1114
weight sum 268 500
precision 1.4741 1.4741
Status
OK
Log

.

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.