EXPERIMENT NO: 4

Aim: To Demonstrate classification rule process on dataset using naive bayes algorithm.

Theory:

Naïve Bayes Classifier Algorithm

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
- It is mainly used in text classification that includes a high-dimensional training dataset.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
- Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

Why is it called Naïve Bayes?

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:

Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.

Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem.

Bayes' Theorem:

- Bayes' theorem is also known as Bayes' Rule or Bayes' law, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

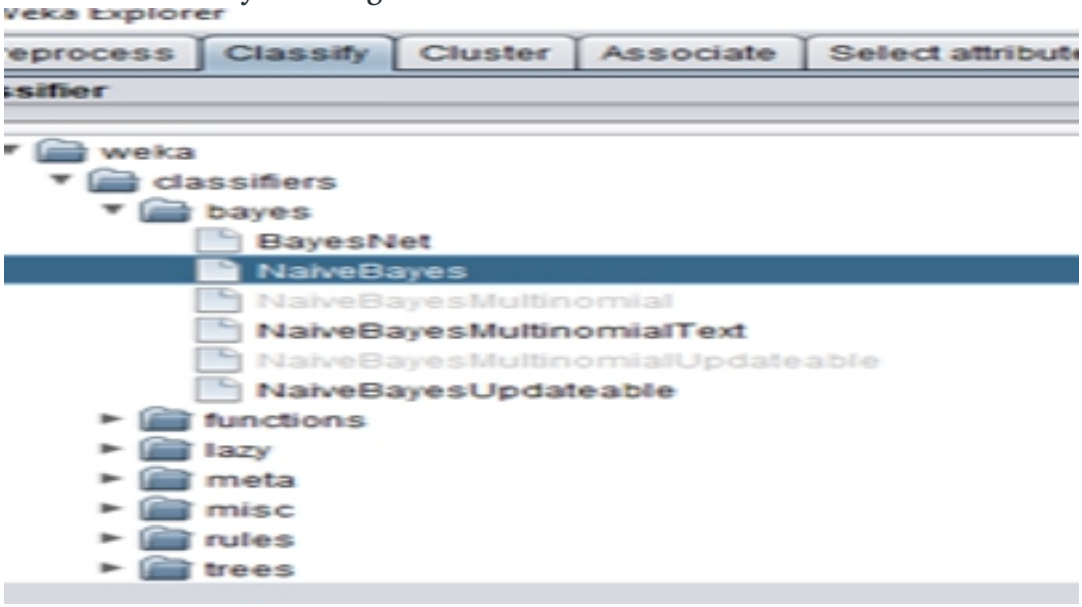
P(B) is Marginal Probability: Probability of Evidence.

Working of Naïve Bayes' Classifier:

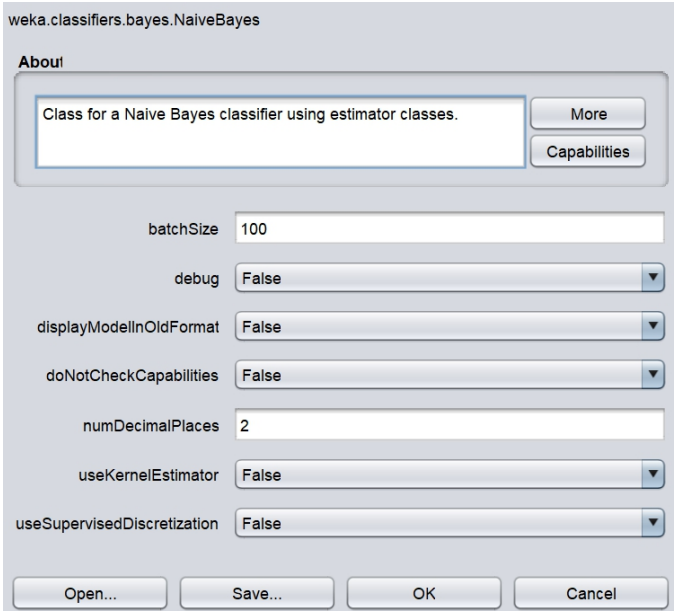
Working of Naïve Bayes' Classifier can be understood with the help of the below example:
Suppose we have a dataset of weather conditions and corresponding target variable "Play".
So using this dataset we need to decide that whether we should play or not on a particular day according to the weather conditions. So to solve this problem, we need to follow the below steps:

- Convert the given dataset into frequency tables.
- Generate Likelihood table by finding the probabilities of given features.
- Now, use Bayes theorem to calculate the posterior probability.

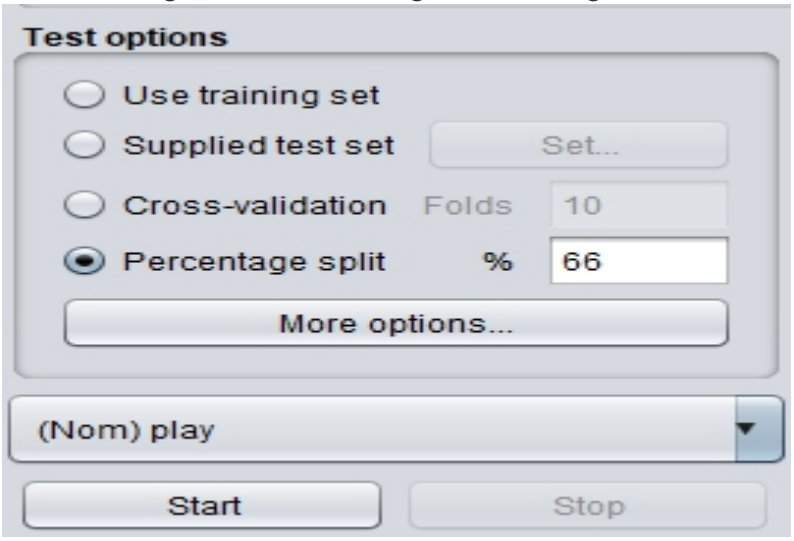
Step 1. Initially, we have to load the required dataset in the weka tool using choose file option. Here we are selecting the weather-nominal dataset to execute.
Step 2. Now we have to go to the classify tab on the top left side and click on the choose button and select the Naive Bayesian algorithm in it.



Step 3. Now to change the parameters click on the right side at the choose button, and we are accepting the default values in this example.



Step 4: We choose the Percentage split as our measurement method from the "Test" choices in the main panel. Since we don't have a separate test data collection, we'll use the percentage split of 66 percent to get a good idea of the model's accuracy. Our dataset contains 14 examples, with h9 being used for training and 5 being used for testing.



Step 5: To generate the model, we now click "start." When the model is done, the evaluation statistic will appear in the right panel.

Output:

```
Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances           3               60      %
Incorrectly Classified Instances         2               40      %
Kappa statistic                          0
Mean absolute error                      0.4437
Root mean squared error                  0.5023
Relative absolute error                 93.8582 %
Root relative squared error            102.2471 %
Total Number of Instances                5

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.600      1.000    0.750      ?        0.667     0.867     yes
                 0.000    0.000    ?          0.000    ?          ?        0.667     0.583     no
Weighted Avg.    0.600    0.600    ?          0.600    ?          ?        0.667     0.753

=== Confusion Matrix ===

 a b   <-- classified as
 3 0 | a = yes
 2 0 | b = no
```

Result: Thus the classification on Data set is performed by NaiveBayes Classification.

Viva Questions.

Q.1) What is a Naïve Bayes Classifier ?

Q.2) What Are the Basic Assumption ?

Q.3) What are the advantages of Naïve Bayes Classifier ?

Q.4) Name the different Problem statement you can solve using Naive Baye's