S. B. JAIN INSTITUTE OF TECHNOLOGY,
MANAGEMENT & RESEARCH, NAGPUR.

Practical No. 6

Aim: Implementation of Naïve Bayes Algorithm for Classification in Data Mining by selecting any available dataset.

Name of Student: Shrutika Pradeep Bagdi

Roll No.: CS22130

Semester/Year: 6th / 3rd

Academic Session: 2024-2025

Date of Performance:

Date of Submission:

AIM: Implementation of Naïve Bayes Algorithm for Classification in Data Mining by selecting any available dataset.

OBJECTIVE/EXPECTED LEARNING OUTCOME:

The objectives and expected learning outcome of this practical are:

- The Naive Bayes classifier is an algorithm used to classify new data instances using a set of known training data.
- It is a good algorithm for classification; however, the number of features must be equal to the number of attributes in the data
- It is computationally expensive when used to classify a large number of items.

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement: Computer System with high configurations

Software Requirement: Weka Tool-3.6.9

THEORY:

Naive Bayes is a popular algorithm used for classification tasks in machine learning. It is based on Bayes' theorem and assumes that the features in the data are independent of each other. Naive Bayes is particularly useful for text classification, where the features are often words that appear in the text.

Classification Rule Process using Naive Bayes Algorithm:

1. **Data Preparation:**

The first step in the classification rule process is to prepare the data. This involves selecting the dataset, cleaning the data, and dividing it into training and testing sets. The data should be cleaned to remove any missing values, outliers, or errors that may affect the performance of the algorithm.

2. **Model Training:**

Once the data is prepared, the next step is to train the Naive Bayes model on the training set. The algorithm calculates the probability of each class given the input features. The model uses the training set to estimate the parameters of the probability distribution function for each class.

3. **Model Evaluation:**

After the model is trained, it is evaluated on the testing set to determine its accuracy. The accuracy of the model is measured using various metrics such as precision, recall, and F1 score. These metrics provide an indication of how well the model is performing and help to identify any areas where it may need improvement.

4. **Making Predictions:**

Once the model is trained and evaluated, it can be used to make predictions on new data. The algorithm calculates the probability of each class given the input features and selects the class with the highest probability as the predicted class.

Types of Naive Bayes Algorithms:

There are three types of Naive Bayes algorithms:

1. **Gaussian Naive Bayes:** This algorithm is used for continuous data and assumes that the features are normally distributed.
2. **Multinomial Naive Bayes**: This algorithm is used for discrete data and assumes that the features are counts.
3. **Bernoulli Naive Bayes:** This algorithm is also used for discrete data but assumes that the features are binary (i.e., they take on values of 0 or 1).

Advantages of Naive Bayes:

1. Naive Bayes is relatively simple to implement and requires fewer computational resources compared to other machine learning algorithms.
2. Naive Bayes can handle a large number of features and is particularly useful for text classification tasks.
3. Naive Bayes is robust to noisy data and can handle missing values.

Disadvantages of Naive Bayes:

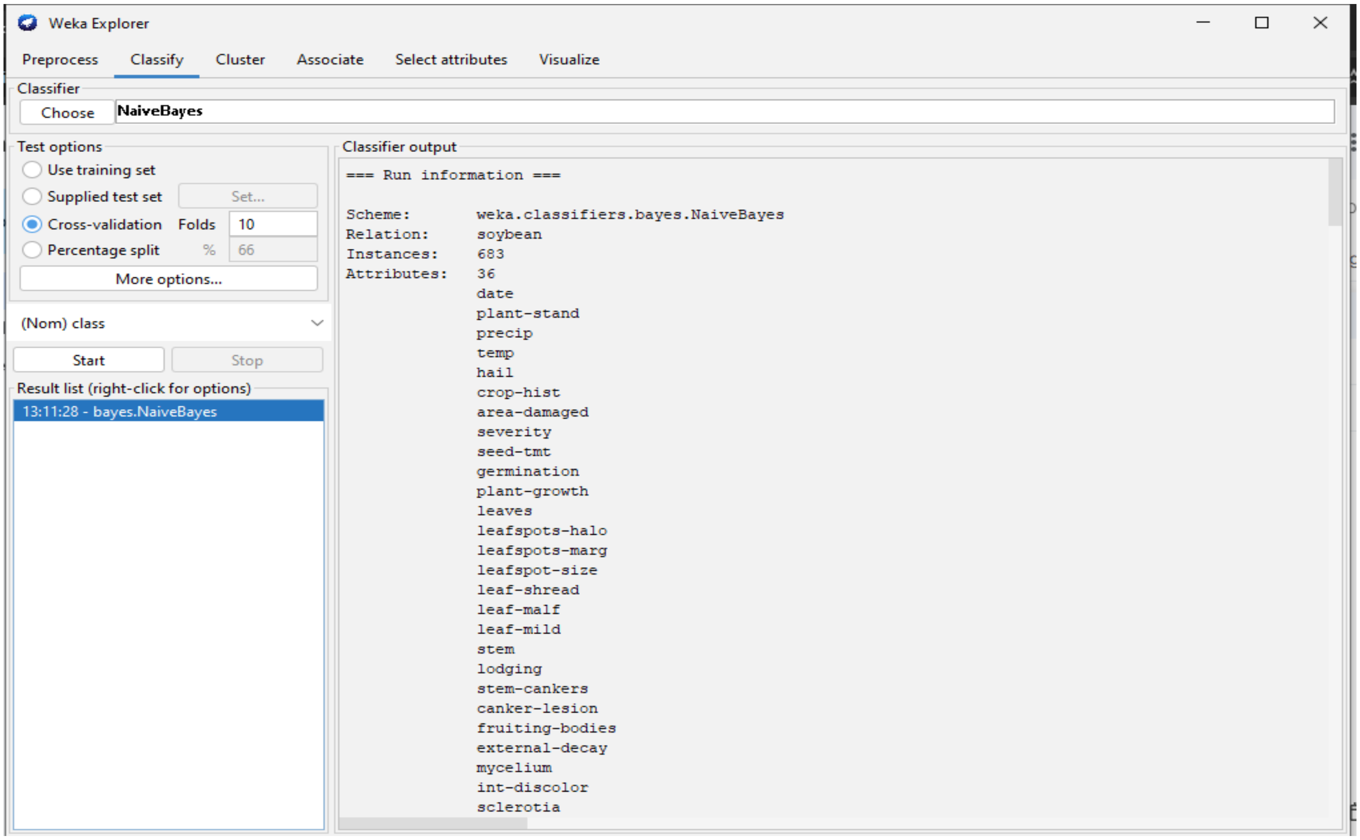
1. Naive Bayes assumes that the features are independent of each other, which may not always be the case in real-world datasets.
2. Naive Bayes can suffer from the problem of zero-frequency, where a feature is not present in the training set, leading to incorrect probabilities.

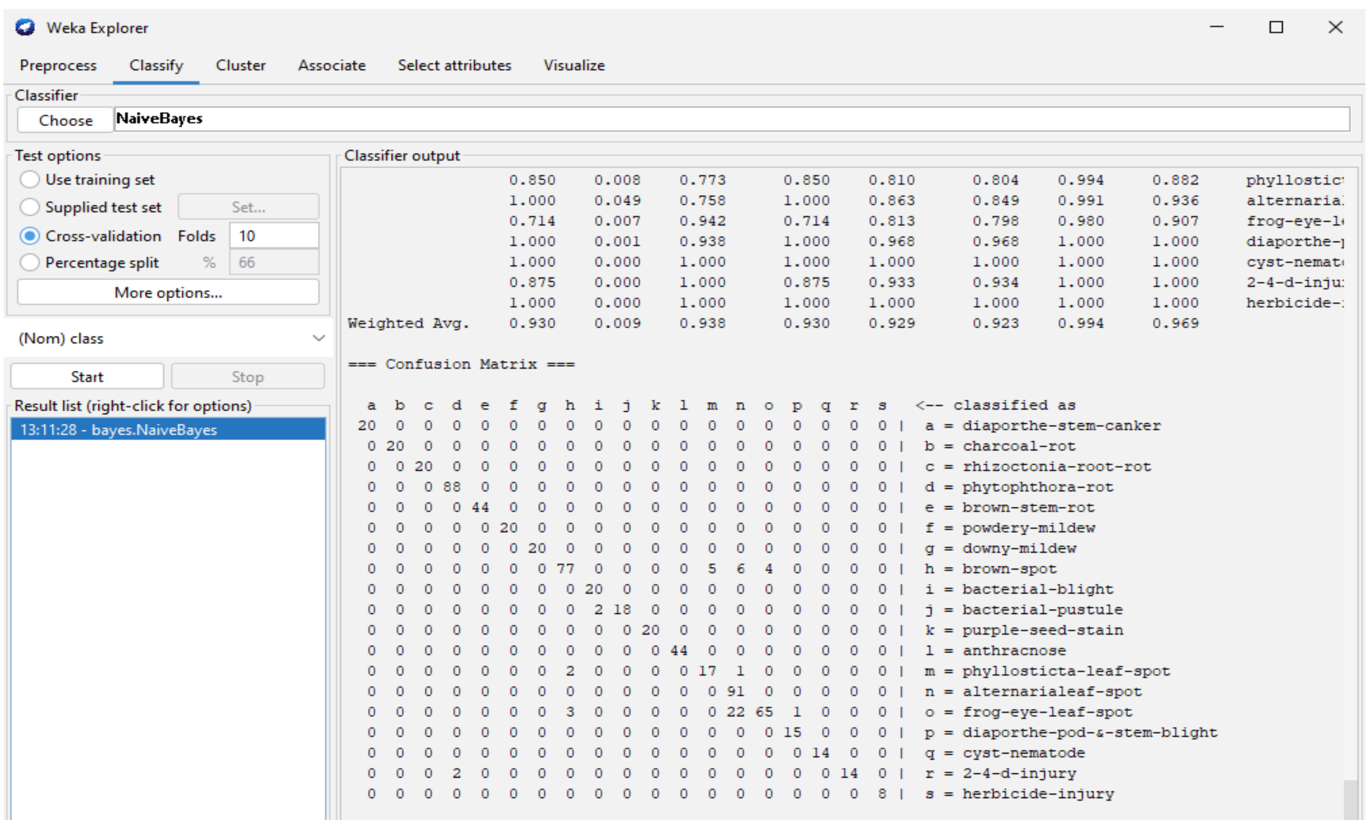
Applications of Naive Bayes:

1. **Text classification:** Naive Bayes is commonly used for text classification tasks, such as spam detection and sentiment analysis.
2. **Medical diagnosis:** Naive Bayes can be used for medical diagnosis tasks, such as identifying diseases based on symptoms.
3. **Fraud detection:** Naive Bayes can be used for fraud detection tasks, such as identifying fraudulent credit card transactions.

Procedure:

OUTPUT (SCREENSHOTS):





CONCLUSION:

DISCUSSION AND VIVA VOCE:

1. How does the algorithm handle missing data?
2. What is the "naive" assumption made in Naive Bayes classification?
3. What is Naive Bayes classification and how does it work?
4. What are some common applications of Naive Bayes classification?
5. What are the limitations of Naive Bayes classification?
6. Can Naive Bayes handle multi-class classification problems?

REFERENCE:

- https://www.google.com/search?q=naivebayes+classiciatiop in+WEKA&sa=X&ved=2ahUKEwjCrY-3vc7-AhVlSWwGHRnJB3sQ1QJ6BAg4EAE&biw=1350&bih=583&dpr=1
- https://www.analyticsvidhya.com/blog/2021/09/naive-bayes-algorithm-a-complete-guide-for-data-science-enthusiasts/

Data Mining – Concepts and Techniques, Jiawei Han & Micheline Kamber, Morgan Kauf-mann Publishers, Elsevier, 2nd Edition, 2006.

Observation book: (3)	Viva-Voce (3)	`Quality of Submission and timely Evaluation (4)
Total:		**Sign with date:**