**Name:** Nikhil Sanjay Shinde

**PRN No:**122B1F120 **Roll No:** TYITB120 **Assignment No. 6  
AIM:** Assignment on Naïve Bayes Classifier **PREREQUISITE:** Python programming

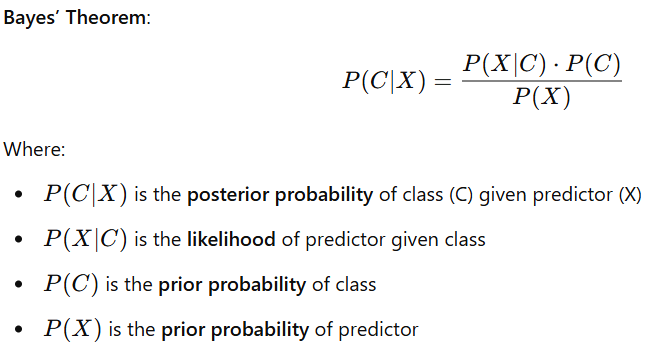
**THEORY:**

Naïve Bayes Classifier is a supervised learning method which uses Bayes' Theorem under independence between predictors. Surprisingly, this "naïve" assumption works well in many complex real-world situations, especially in text classification tasks.

It relies on the fact that the probability of a class based on a set of features can be calculated from the prior probabilities of the class and the probability of features given the class. Naïve Bayes algorithm is very efficient and effective for high-dimensional large datasets.

**Concept of Naïve Bayes Classifier:**

To see how Naïve Bayes is applied, consider predicting whether an email is spam or not spam given words in the email. If words like "discount" or "free" occur regularly in spam messages, then looking for those words raises the chance a new message is spam.



The classifier selects the class label with the **highest posterior probability**.

**Working Mechanism of Naïve Bayes Classifier:**

* Calculate Prior Probability
* Frequency of every class in the training set.
* Calculate Likelihood
* Conditional probability of every feature given every class.
* Apply Bayes' Theorem
* Add prior and likelihood to calculate posterior probabilities.
* Predict Class
* Select the class with the highest posterior probability.

**Applications of Naïve Bayes Classifier:**

* **Spam Filtering:** Identifies if an email is spam using keywords.
* **Sentiment Analysis:** Examines text to identify positive, negative, or neutral sentiment.
* **Document Categorization:** Categorizes articles or news into topics.
* **Medical Diagnosis:** Makes predictions of diseases based on symptoms and history.
* **Recommendation Systems:** Recommends products based on user behavior.

**Advantages of Naïve Bayes Classifier:**

* **Easy and Quick**: Easy to implement and needs less training data.
* **Efficient with Large Datasets:** Most effective for text classification problems.
* **Works Well with Noisy Data:** Insensitive to irrelevant features.
* **Operates with Categorical and Numerical Data:** Supports multiple data types.

**Naïve Bayes Classifier's Disadvantages:**

* **Strong Independence Assumption:** Imposes that all features are independent, which is seldom true with real-world data.
* **Zero Frequency Problem:** Gives zero probability to unseen features if smoothing is not used.
* **Not Suitable for Continuous Variables:** Needs Gaussian or other assumptions for numerical data.

**CONCLUSION**

Naïve Bayes Classifier is a widely used and efficient classification algorithm employed across vast industries. Its speed and simplicity render it the first option for text mining, spam detection, and classification tasks despite the violation of the independence assumption. With proper preprocessing and management, Naïve Bayes is a top choice algorithm for rapid and precise predictions.