Name: Bhavya Bhushan Wade

Subject: Data Warehouse Mining

Roll No.: 77

Sem: 5

**Topic: Naïve Bayes**

Naive Bayes is a probabilistic algorithm widely used in classification tasks within Data Warehouse Mining. It's based on Bayes' theorem, which calculates the probability of an event occurring based on prior knowledge of related events. In the context of data mining, Naive Bayes is primarily used for classification tasks, where it predicts the class or category of an instance based on the features associated with it.

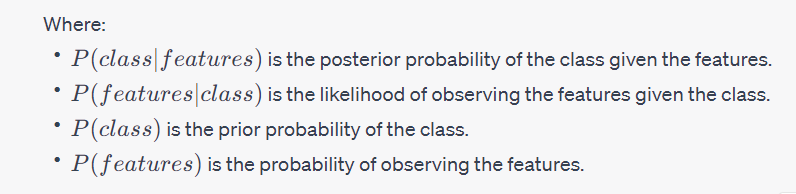
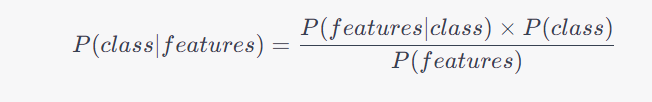
How Naive Bayes works:

Assumption of Independence:

Naive Bayes operates on the assumption that the features used to describe an instance are independent of each other. This is a simplifying assumption to make the calculations more manageable.

Bayes' Theorem:

Naive Bayes uses Bayes' theorem to calculate the posterior probability of a class given the features of an instance. The formula is:



P(features) is the probability of observing the features.

Naive Bayes is efficient, easy to implement, and works well for text classification, spam filtering, sentiment analysis, and many other classification tasks.

**Conclusion:**

Naive Bayes is a powerful probabilistic classification algorithm. Its simplicity, efficiency, and effectiveness in various applications make it a fundamental tool in the data mining toolkit. It operates under the assumption of feature independence, allowing it to calculate probabilities and make predictions efficiently. By employing Naive Bayes, data analysts can classify data into distinct categories or classes, making it an invaluable asset for tasks such as spam detection, sentiment analysis, and medical diagnosis.