

Naïve Bayes Practical Viva Q&A

1. Q: What is the Naïve Bayes algorithm?
A: It's a classification algorithm based on Bayes' Theorem that assumes independence between predictors.
2. Q: Why is it called “naïve”?
A: Because it assumes that all features are independent, which is a simplification of reality.
3. Q: What type of problems can Naïve Bayes solve?
A: Classification problems like spam detection, sentiment analysis, and species classification.
4. Q: What dataset did we use in this experiment?
A: The Iris dataset, which contains features of iris flowers and their species.
5. Q: What are the target classes in the Iris dataset?
A: Setosa, Versicolor, and Virginica.
6. Q: What is the role of LabelEncoder?
A: It converts categorical class labels into numeric values required by the model.
7. Q: What does **GaussianNB** assume about the data?
A: It assumes that the features follow a Gaussian (normal) distribution.
8. Q: Why do we split the dataset?
A: To train the model on one part and test its performance on unseen data.
9. Q: What is a confusion matrix?
A: A matrix that shows the number of correct and incorrect predictions for each class.
10. Q: What are TP, FP, TN, and FN?
A: True Positives, False Positives, True Negatives, and False Negatives — used to calculate performance metrics.
11. Q: Define accuracy.
A: The ratio of correctly predicted observations to the total

observations.

12.Q: Define precision.

A: The ratio of correctly predicted positive observations to the total predicted positives.

13.Q: Define recall (sensitivity).

A: The ratio of correctly predicted positives to all actual positives.

14.Q: Define error rate.

A: The proportion of incorrect predictions out of total predictions.

15.Q: How do you interpret a high recall and low precision?

A: The model captures most positives but also produces many false positives.

16.Q: What does F1-score represent?

A: It is the harmonic mean of precision and recall, providing a balanced measure.

17.Q: Why use Gaussian Naïve Bayes over other classifiers?

A: It's simple, fast, and performs well with small datasets and normally distributed features.

18.Q: What are the advantages of Naïve Bayes?

A: It's fast, requires less training data, and handles high-dimensional data well.

19.Q: What are the disadvantages of Naïve Bayes?

A: The strong assumption of feature independence may reduce accuracy in some cases.

20.Q: How many features are used from the Iris dataset?

A: Four: Sepal length, Sepal width, Petal length, Petal width.

21.Q: Can Naïve Bayes be used for multi-class classification?

A: Yes, it naturally supports multi-class classification.

22.Q: Why do we visualize the confusion matrix?

A: To easily interpret which classes are being correctly or incorrectly predicted.

23.Q: What does `.fit()` do in sklearn?

A: It trains the model on the provided training data.

24.Q: What does `.predict()` do in sklearn?

A: It predicts the class labels for the test data based on the trained model.

25.Q: What is the role of `random_state` in `train_test_split()`?

A: It ensures reproducibility by setting the seed for random splitting.