**NAÏVE BAYES CLASSIFIER**

**Ex no: 4  
RUBAN S**

**Aim:**

To build a Naive Bayes classifier to predict the class of a given dataset using the provided

features.

**What is Naive Bayes?**

Naive Bayes is a family of probabilistic machine learning algorithms based on Bayes

Theorem with an assumption of independence among the features. The Naive Bayes classifier

assumes that the presence of a feature in a class is not related to any other feature. Naïve

Bayes is a classification algorithm for binary and multi-class classification problems.

**Procedure**

1. **Prepare the data**:
   * Encode categorical variables into numerical values if necessary.
   * Split the data into features (X) and target variable (y).
   * Divide the data into training and testing sets.
2. **Create and train the initial Naive Bayes model**:
   * Choose an appropriate Naive Bayes variant (GaussianNB, MultinomialNB, BernoulliNB).
   * Train the model on the training data.
3. **Evaluate the initial model**:
   * Make predictions on the test data.
   * Print classification report and confusion matrix.
   * Print the accuracy of the model.
4. **Visualize the feature correlations**:
   * Calculate the correlation matrix of the features and the target variable.
   * Plot the correlation matrix using a heatmap.
5. **Select features based on correlation**:
   * Identify the feature with the highest correlation with the target variable.
   * Select features with low correlation to the highest correlation feature.
6. **Tune the model**:
   * Use techniques like cross-validation to find the best model and hyperparameters.
   * Print the best parameters and their score.
7. **Evaluate the improved model**:
   * Make predictions on the test data.
   * Print classification report and confusion matrix.
   * Print the accuracy of the improved model.
8. **Visualize the results**:
   * Plot the accuracy scores of the models.
   * Plot confusion matrices for the models.
9. **Perform k-fold cross-validation**:
   * Evaluate models using k-fold cross-validation.
   * Print the cross-validated accuracy and standard deviation for each model.
   * Plot cross-validation results.

**Choosing Features in Naive Bayes**

Features are selected based on their relevance to the classification task and their statistical

properties. In this exercise, features are chosen based on their correlation with the target

variable and their ability to contribute to the prediction.

**When Can Overfitting Occur?**

Overfitting can occur if the model is too complex or if there are too many features relative to

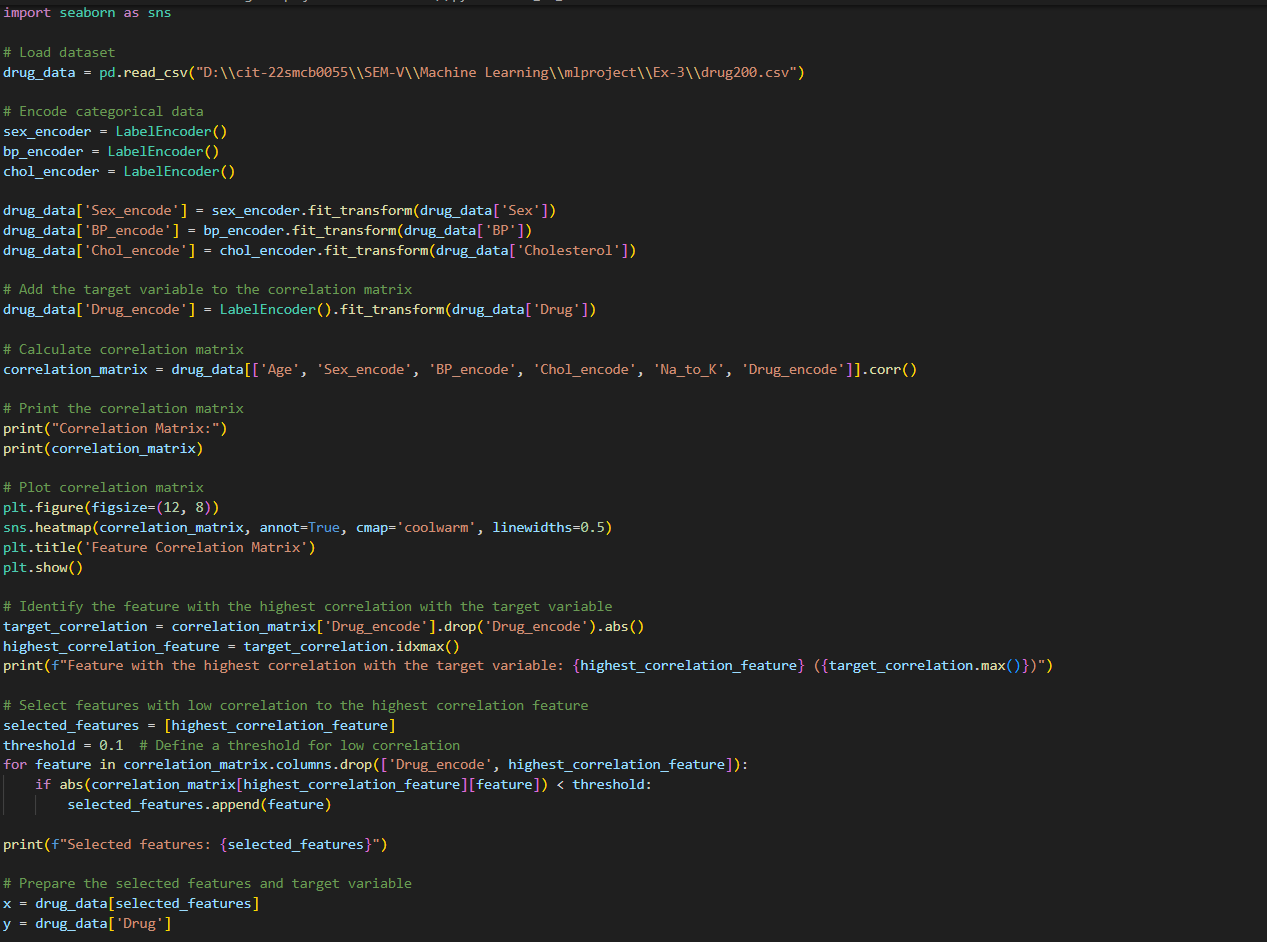
the number of training examples. It can also happen if the training data is not representative

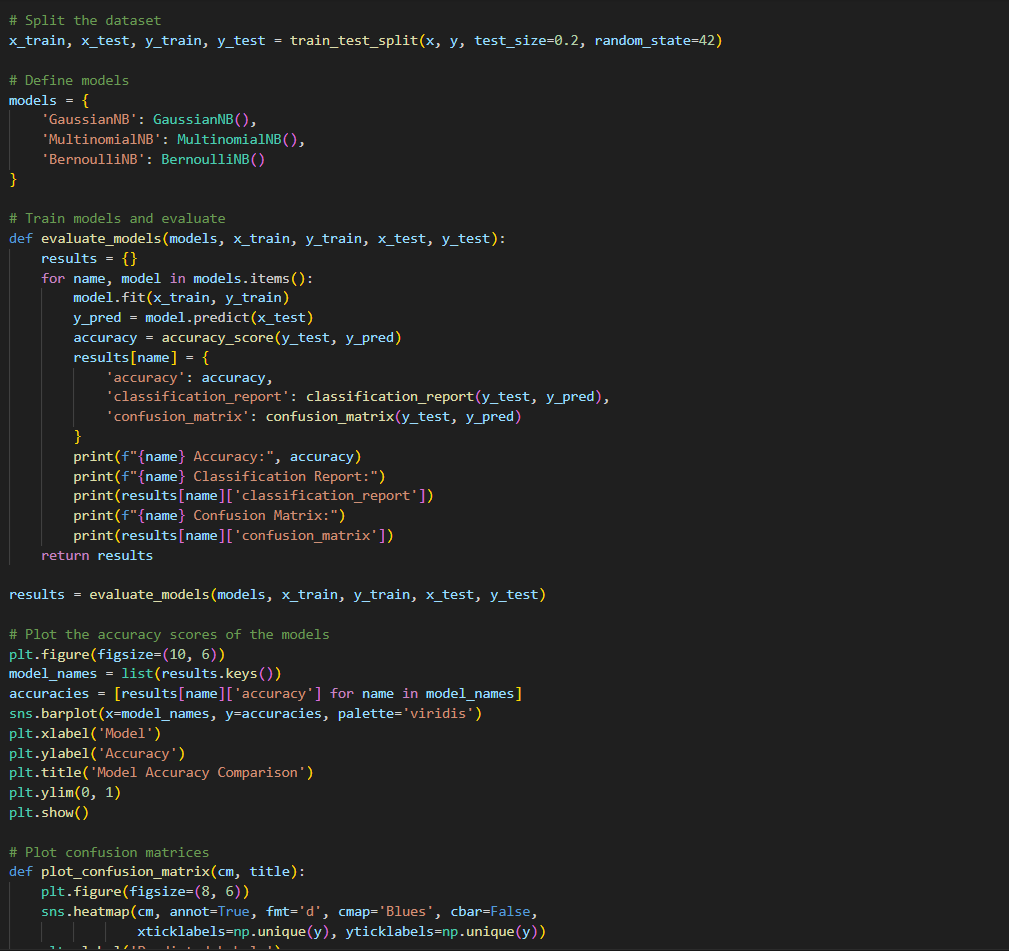
of the overall population. Naive Bayes is generally less prone to overfitting due to its

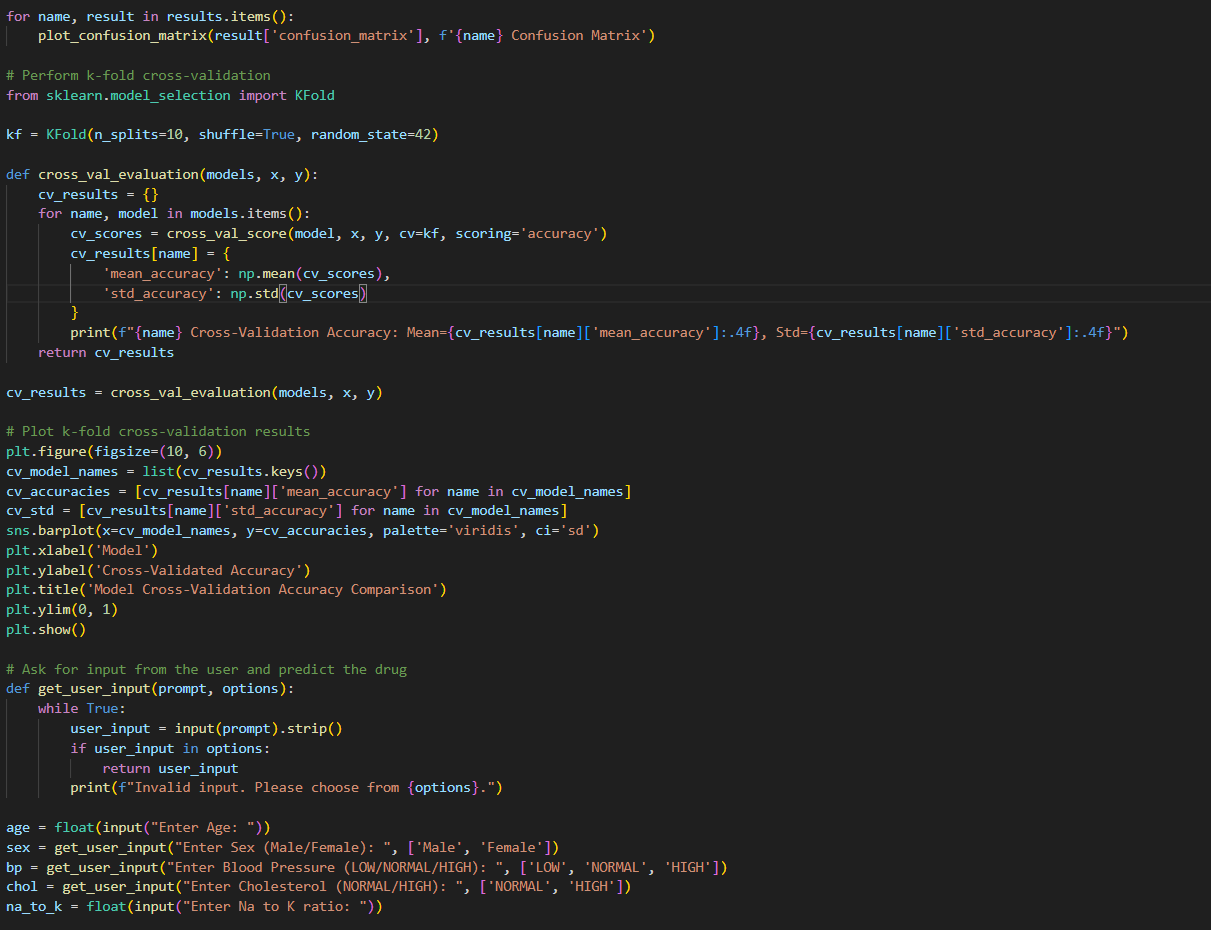
simplicity, but it can still occur if the assumptions of independence among features do not

hold.

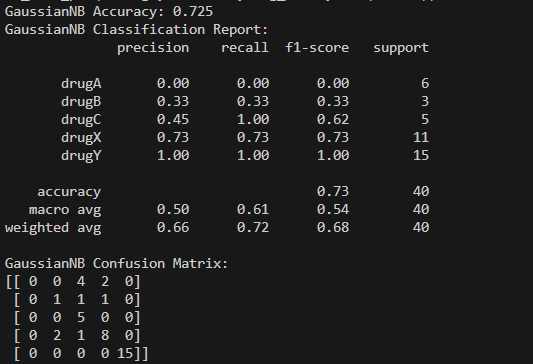
**CODE:**

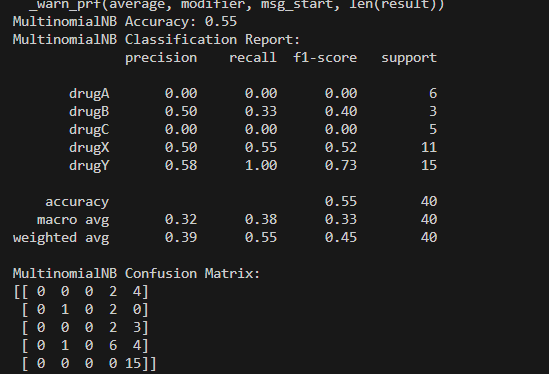
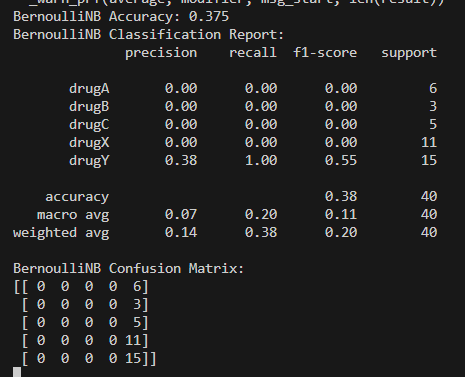
****

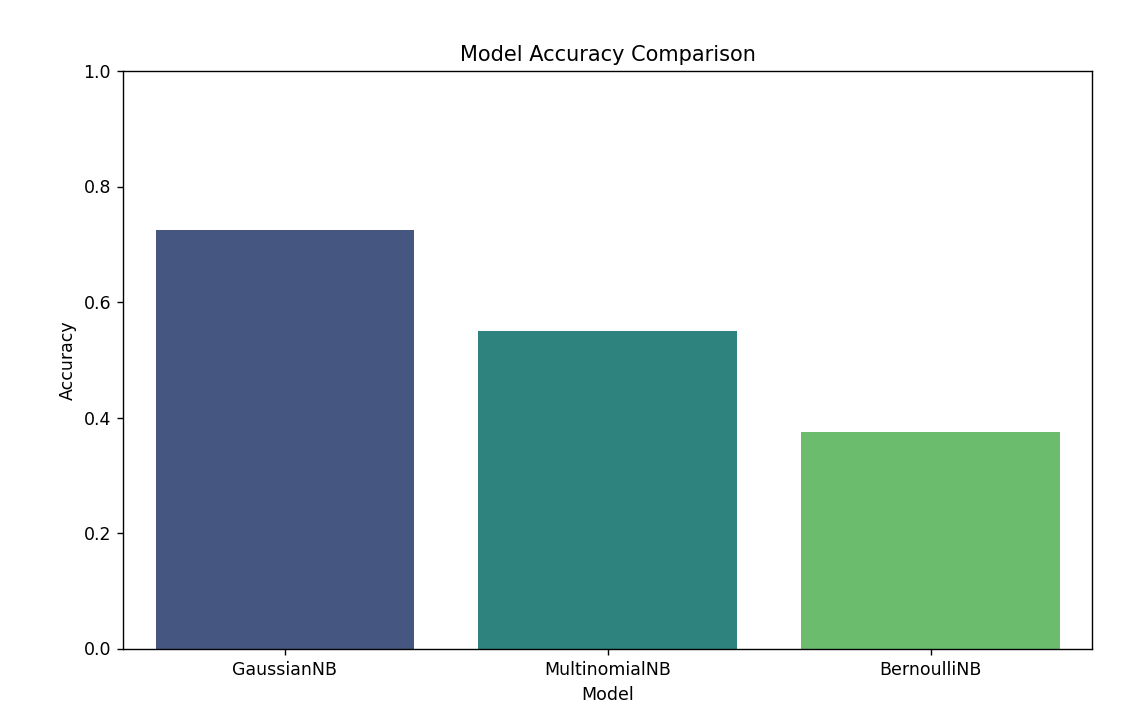
****

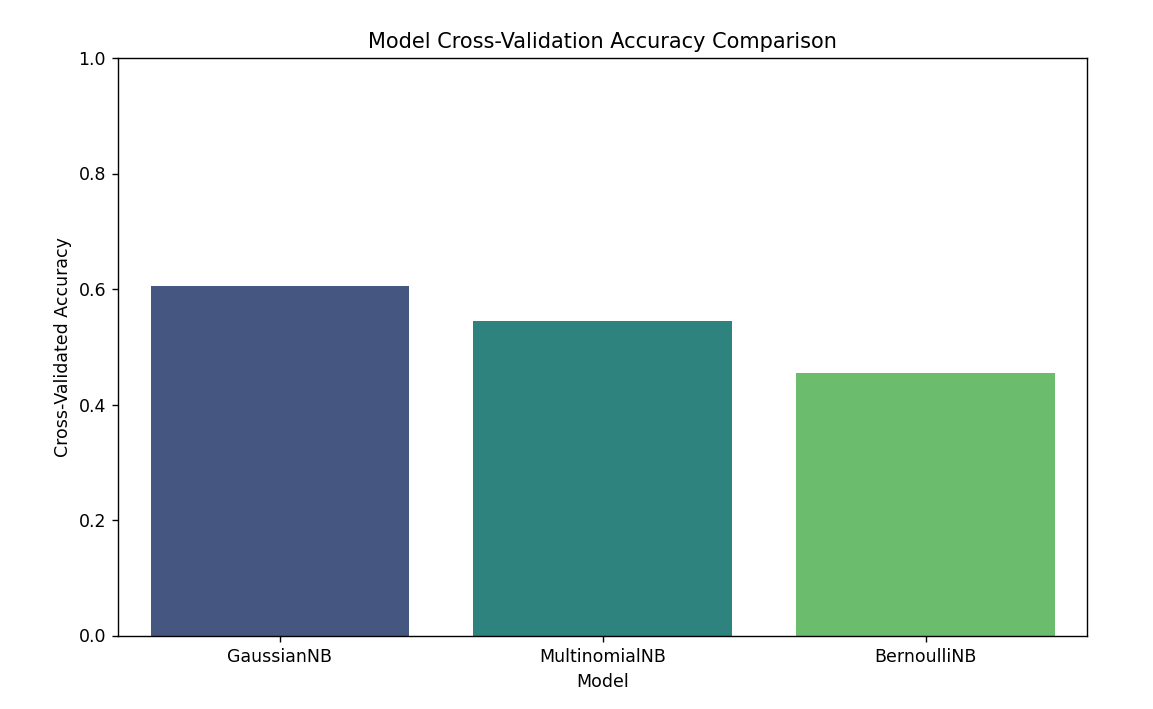
****

**Output**

****

**** ****

****

****

**Conclusion**

Naive Bayes classifiers are a straightforward and powerful tool for classification tasks. They

are particularly useful when the assumption of feature independence holds, and they provide

a simple yet effective way to make predictions based on the probability of different

outcomes.