**DECISION TREE CLASSIFIER Ex no: 3**

**RUBAN S**

**What are decision trees?**

Decision trees are a popular machine learning algorithm used for both

classification and regression tasks. They model decisions based on the features of

the data and their outcomes.

## AIM:

To build a decision tree classifier to predict drug types using the provided

dataset.

## Procedure

1. Import necessary libraries and load the dataset
2. Prepare the data:
   * Encode categorical variables (Sex, BP, Cholesterol) into numerical values
   * Split the data into features (X) and target variable (y)
   * Divide the data into training and testing sets
3. Create and train an initial decision tree model:
   * Set a fixed max\_depth of 2
   * Train the model on the training data
4. Evaluate the initial model:
   * Make predictions on the test data
   * Print classification report and confusion matrix
5. Tune the model:
   * Use GridSearchCV to find the best max\_depth
   * Try max\_depth values from 1 to 19
   * Print the best max\_depth and its score
6. Create and train an improved model:
   * Use the best max\_depth found during tuning
   * Train the new model on the training data
7. Evaluate the improved model:
   * Make predictions on the test data
   * Print classification report and confusion matrix
8. Visualize the results:
   * Show the decision tree before tuning
   * Show the decision tree after tuning
   * Plot cross-validation scores for different max\_depth values
9. Create a simple prediction tool:
   * Ask the user to input patient information
   * Use the tuned model to predict the drug type

## Choosing Features in Decision Trees

Features are selected based on their ability to split the data to reduce impurity

(measured by Gini impurity or entropy). In this exercise, we choose features

like Age, Sex, BP, Cholesterol, and Na\_to\_K based on their relevance to

predicting drug types.

## When Can Overfitting Occur?

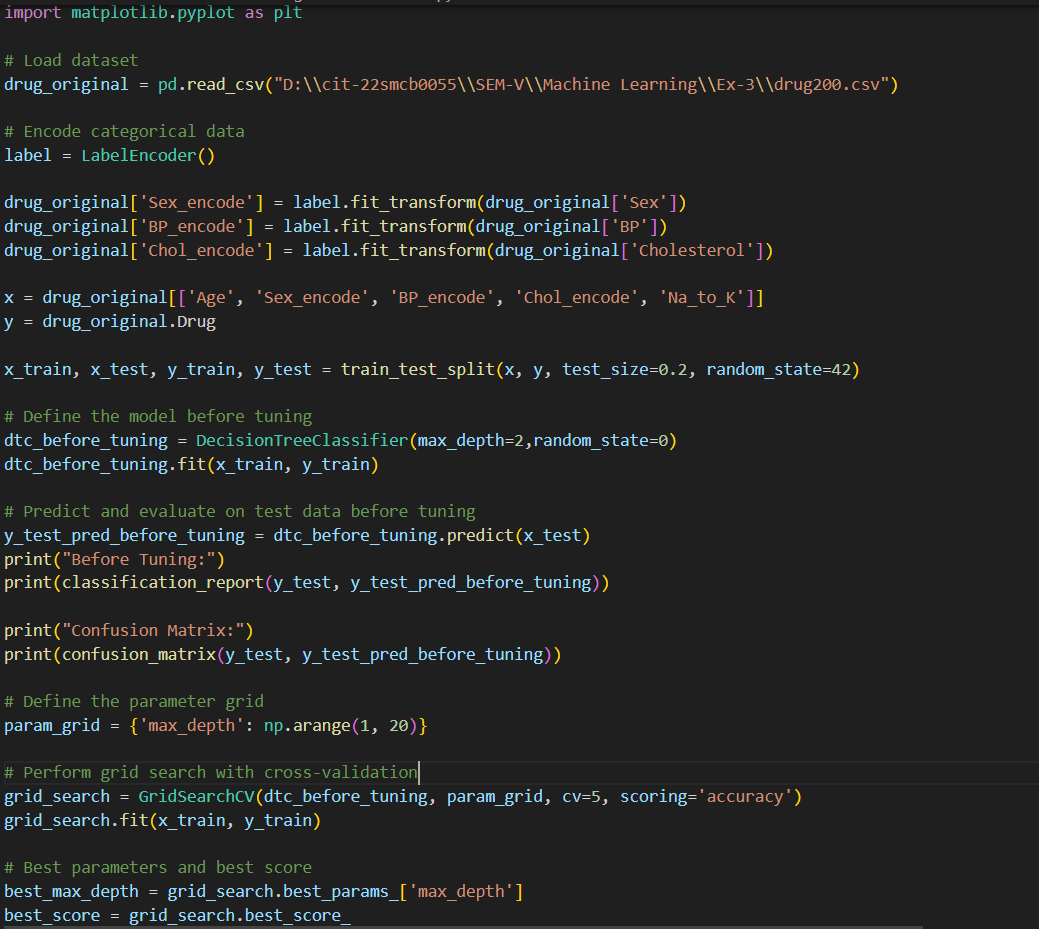
However, as a tree grows in size, it becomes increasingly difficult to

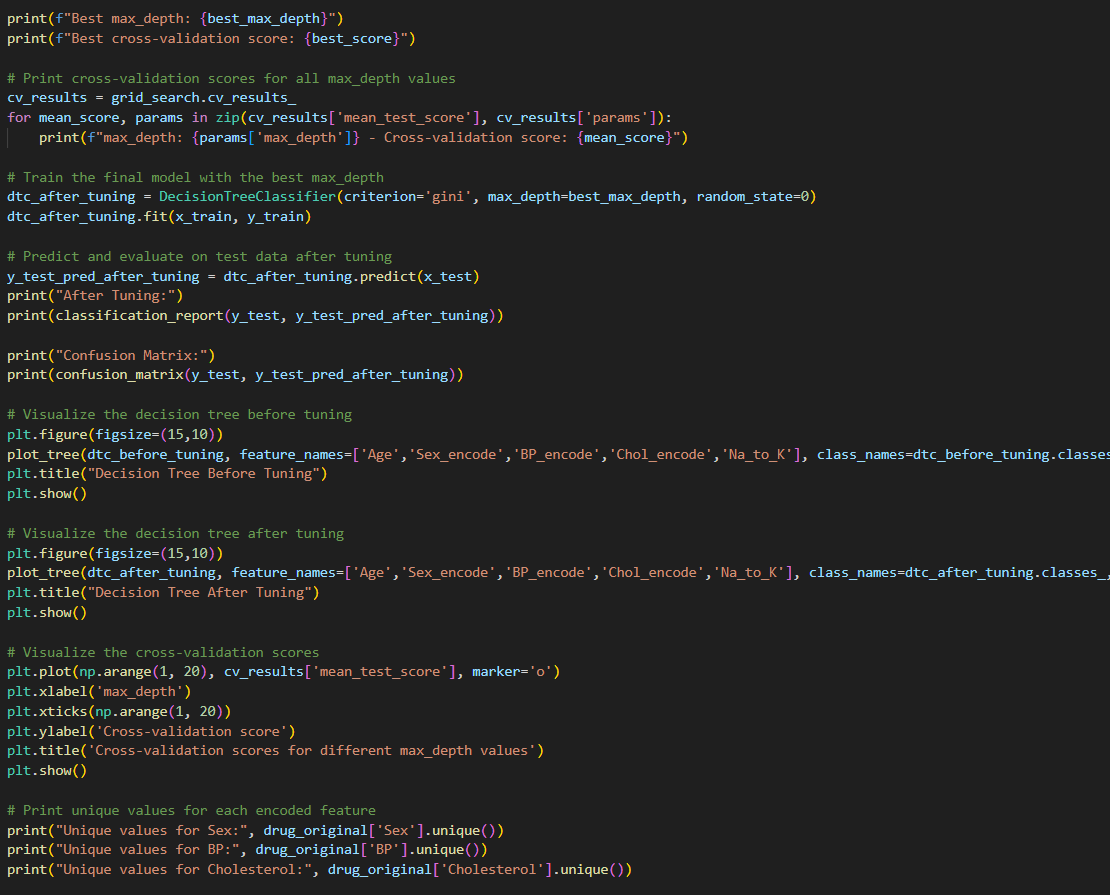
maintain this purity, and it usually results in too little data falling within a

given subtree. When this occurs, it is known as data fragmentation, and it

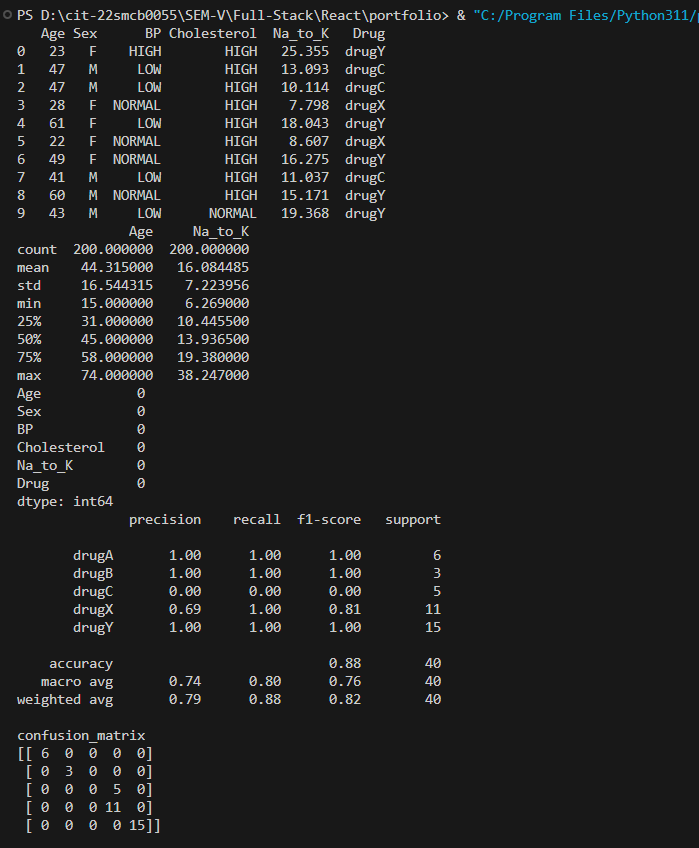
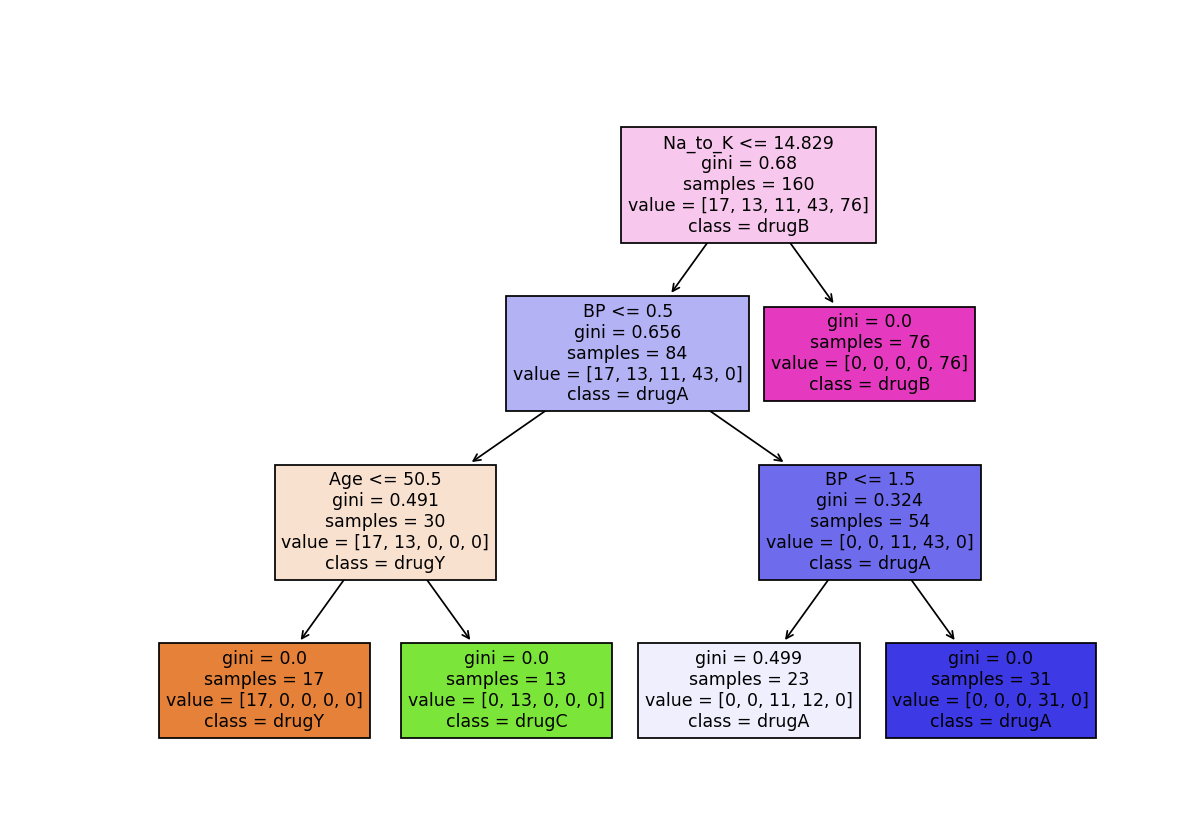
can often lead to overfitting.

## Code





## Output



## Confusion Matrix

A confusion matrix is a table used to evaluate the performance of a

classification algorithm. It shows the true positive, true negative, false positive,

and false negative predictions, providing a comprehensive view of how well the

classifier is performing.

## Conclusion

Decision trees are a straight forward and powerful tool for classification tasks.

They allow us to visualize and interpret the decision-making process, making

them useful for understanding how different features contribute to the

predictions.