**Assignment 3**

**Wine Classification using Decision Tree**

**Name: Saurabh Isane**

**Roll No: 122B1F035**

**Dataset Used: Wine Dataset from sklearn.datasets**

**Objective:**

To understand the concept of Decision Tree Classification and implement it on the Wine dataset to predict the type of wine based on its chemical properties.

**What is a Decision Tree?**

A Decision Tree is a supervised machine learning algorithm used for classification and regression tasks. It splits the data into branches based on feature values, forming a tree-like structure. Each internal node represents a decision rule on a feature, each leaf node represents an output class, and branches represent the conjunction of features that lead to those classes.

**Key Concepts:**

* **Root Node:** The first decision node.
* **Leaf Node:** Final output node with a class label.
* **Splitting:** Dividing data into subsets.
* **Gini Index / Entropy:** Measures of impurity used for deciding splits.
* **Overfitting:** When the model is too complex and fits noise in the training data.

**Steps Performed in the Assignment**

**1. Importing Required Libraries:**

Basic Python libraries like pandas, numpy, matplotlib, seaborn, and sklearn modules were used for modeling and visualization.

**2. Loading the Dataset:**

We used the load\_wine() method from sklearn to import the Wine dataset. The dataset contains 13 features like alcohol, malic acid, ash, etc., and 3 classes of wine.

**3. Data Exploration:**

* Displayed first few rows using head()
* Checked dataset shape and basic statistics using describe()
* Verified null values to ensure data quality

**4. Visualizing the Dataset:**

* Countplot to see class distribution
* Pairplot to understand feature relationships visually

**5. Splitting the Data:**

The dataset was split into training (70%) and testing (30%) sets using train\_test\_split().

**6. Model Building:**

* Used DecisionTreeClassifier() from sklearn
* Fitted the model using fit() on training data

**7. Prediction and Evaluation:**

* Predicted target values for the test set
* Evaluated using accuracy\_score, confusion\_matrix, and classification\_report

**8. Decision Tree Visualization:**

* Used plot\_tree() to visualize the entire decision-making process of the model.

**Results:**

* The model achieved high accuracy in predicting the type of wine.
* Confusion matrix and classification report helped identify the model's performance per class.

**Conclusion:**

Decision Trees are powerful and interpretable models suitable for classification problems. They are easy to visualize and understand but can overfit if not properly controlled. The experiment demonstrates how a Decision Tree can be effectively used to classify wine types using scikit-learn.

**Future Scope:**

* Use cross-validation to avoid overfitting.
* Compare with other models like Random Forest, SVM.
* Perform hyperparameter tuning for better accuracy.