Practical No. 6

Aim :- To implement the decision tree classification algorithm using Iris dataset & visualize the decision tree.

Input :- 1) Dataset : Iris dataset with 4 features : sepal length, sepal width, petal length petal width
2) Target Variable species

output :- Accuracy : 1.0

Theory :-

A decision tree is a supervised machine learning algorithm that is commonly used for classification & regression tasks. It is a tree structured model consisting ~~sons~~ of internal nodes, branches, & leaf nodes. The tree helps in predicting the value of a target variable by learning simple decision rules inferred from data features.

Working principle :

The decision tree stats from the root nodes split the data on the features that results best split the one that increases the purity resulting subsets. The process continues rely until:

1) All samples in a nodes belong to the same
2) No further meaningful splits can be made

Splitting criteria:

To decide the best feature to split, the tree

1) Information Gain (based on Entropy):

Information Gain = Entropy (parent).

2) Gini Index:

$$Gini = 1 - \sum (p_i)^2$$

Both aim to create child nodes that are as pure as possible. Types of Decision Tree:

Classification Tree: output is class label
Regression Tree: output is concinuous value.

Algorithm:

[Step 1]: load the dataset
[Step 2]: split the data into training & testing
[Step 3]: Train the decision Tree classifier on the training data.
[Step 4]: Predict the class labels on test data
[Step 5]: Visualize the tree & evaluate model accu

Conclusion :

The decision tree algorithm was successful implemented using the iris dataset. It was able to classify flower species with high accuracy.
The tree structure provides clear insight into how decisions are made based on feature values.