



Experiment No:5

Course Outcome: CO2

Blooms Level: L3

Aim: To implement Decision Tree Classifier

Abstract:

Decision Tree Classifiers are widely used in machine learning for classification tasks due to their interpretability and simplicity. This project implements a Decision Tree Classifier using the Gini Index as the splitting criterion to predict categorical outcomes from a dataset (e.g., Wine or Breast Cancer). The Gini Index helps select the best feature at each node by measuring impurity. The model is evaluated using accuracy, precision, recall, and F1-score, and the tree structure is visualized for interpretability. The approach demonstrates an effective and easily interpretable method for classification tasks in real-world datasets.

Sample Input and Output:

Case 1:

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa
13	4.8	3	1.4	0.1	Iris-setosa
14	4.3	3	1.1	0.1	Iris-setosa
15	5.8	4	1.2	0.2	Iris-setosa
16	5.7	4.4	1.5	0.4	Iris-setosa
17	5.4	3.9	1.3	0.4	Iris-setosa
18	5.1	3.5	1.4	0.3	Iris-setosa



Sample Output

Accuracy: 1.00

Case 3:

Theory:

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. **It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.**
- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
- The decisions or the test are performed on the basis of features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we can use the CART algorithm, which stands for Classification and Regression Tree algorithm, ID3 Algorithm, C4.5 etc.
- A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.
- Note: A decision tree can contain categorical data (YES/NO) as well as numeric data.

Program:

Dataset Information

The Iris dataset consists of 150 samples of iris flowers, 3 classes (Setosa, Versicolor, and Virginica) of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2. The latter are not linearly separable from each other.

Attribute information:

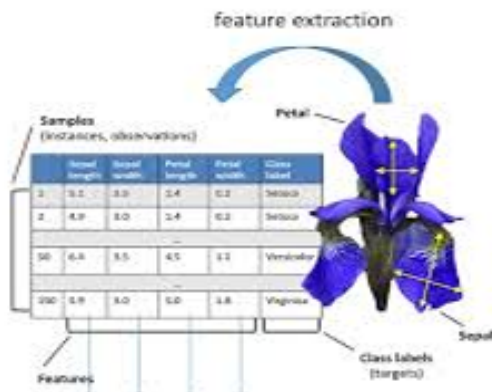
1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm.
5. class:-Iris Setosa--Iris Versicolour—Iris Virginica



Iris Versicolor

Iris Setosa

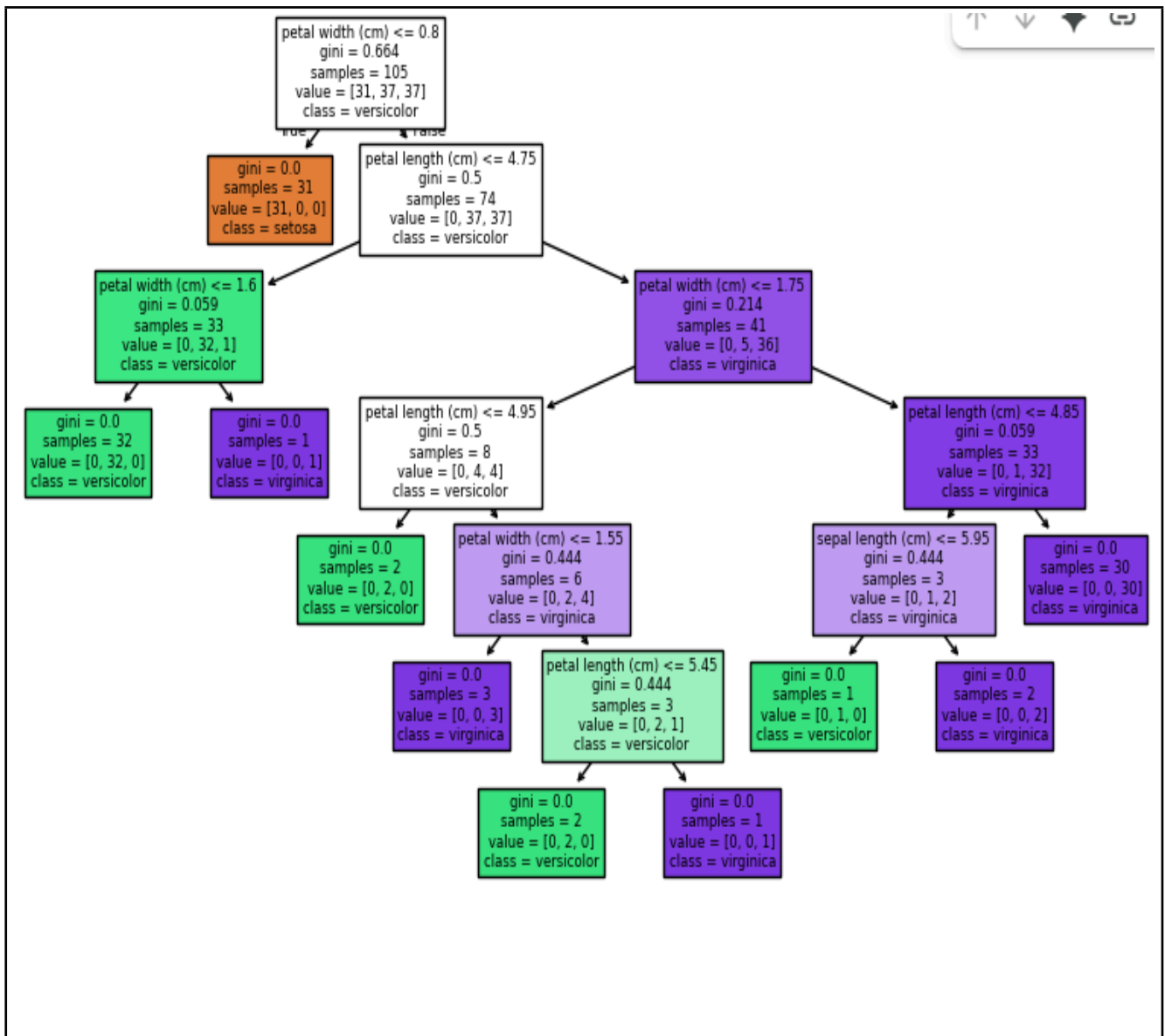
Iris Virginica



Output:



```
1 # import necessary libraries
2 from sklearn import datasets
3 from sklearn.model_selection import train_test_split
4 from sklearn.tree import DecisionTreeClassifier, plot_tree
5 import matplotlib.pyplot as plt
6
7 # load the iris dataset
8 iris = datasets.load_iris()
9 X = iris.data          # features
10 y = iris.target        # labels
11
12 # split into train and test sets
13 X_train, X_test, y_train, y_test = train_test_split(
14     X, y, test_size=0.3, random_state=42)
15
16 # initialize and train the decision tree classifier
17 clf = DecisionTreeClassifier()
18 clf.fit(X_train, y_train)
19
20 # evaluate the model
21 accuracy = clf.score(X_test, y_test)
22 print(f"Accuracy: {accuracy:.2f}")
23
24 # visualize the tree
25 plt.figure(figsize=(10,6))
26 plot_tree(clf, feature_names=iris.feature_names, class_names=iris.target_names, filled=True)
27 plt.show()
28
```



Conclusion: Different decision tree algorithms are ID3, C4.5 and CART. Thus, in this experiment we have implemented Gini Index algorithm which uses information gain attribute selection method and successfully classified the iris data set using python.

Exercise 1:

Breast cancer diagnosis is critical for timely treatment and patient survival. Using the Breast Cancer Wisconsin dataset, the goal is to develop a Decision Tree Classifier that accurately predicts whether a tumor is benign or malignant based on cell nuclei features, and evaluate its performance for reliable medical decision support.

Students shall draw flowchart of exercise question in the writeup and submit.

Exercise 2:



How can we utilize a Decision Tree Classifier to accurately predict the type of wine from three different cultivars based on their chemical properties, and how can we visualize the resulting tree structure to interpret the decision rules behind the classification?

Students shall draw flowchart of exercise question in the writeup and submit.