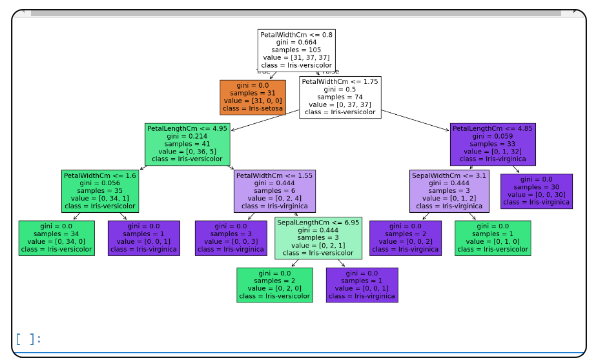
Decision Tree Classifier

Goal: Train and evaluate a Decision Tree Classifier using the Iris dataset to predict the species of iris flowers based on their features.

Action:

1. We overviewed the dataset to get an understanding of the missing values and other related errors in the dataset. We also obtained the summary statistics of the data.
2. Selected the features and target variable. The columns chosen as input features were ‘SepalLengthCm’, ‘SepalWidthCm’, etc. and the column we want to predict is ‘Species’.
3. Created the input features (x) and target variable (y) from the data. Also made sure that the missing values are filled with 0 ensuring no gaps in the data.
4. Split the data into two parts; 70% training and 30% testing set.
5. The evaluation of accuracy of the model was done by comparing the predicted values and the actual values. Also used confusion matrix to understand how many predictions were correct.
6. Finally, we plotted the decision tree to visually understand how the model make decisions.



Output:

Nodes: Each box in the tree represents a node. It shows the feature used for splitting (e.g., PetalWidthCm), the threshold value (e.g., <= 0.8), the Gini impurity, the number of samples at that node, the value distribution of the samples (e.g., [31, 37, 37]), and the predicted class.

Branches: The lines connecting the nodes show the decision path based on the feature threshold. For example, if PetalWidthCm <= 0.8, it goes to the left branch, otherwise, to the right.

Leaves: The terminal nodes (leaves) represent the final prediction. The color indicates the class, and the node details show the majority class in that subset.