

Experiment No. 4

Aim: To implement Decision Tree classifier models to perform supervised classification and evaluate model performance.

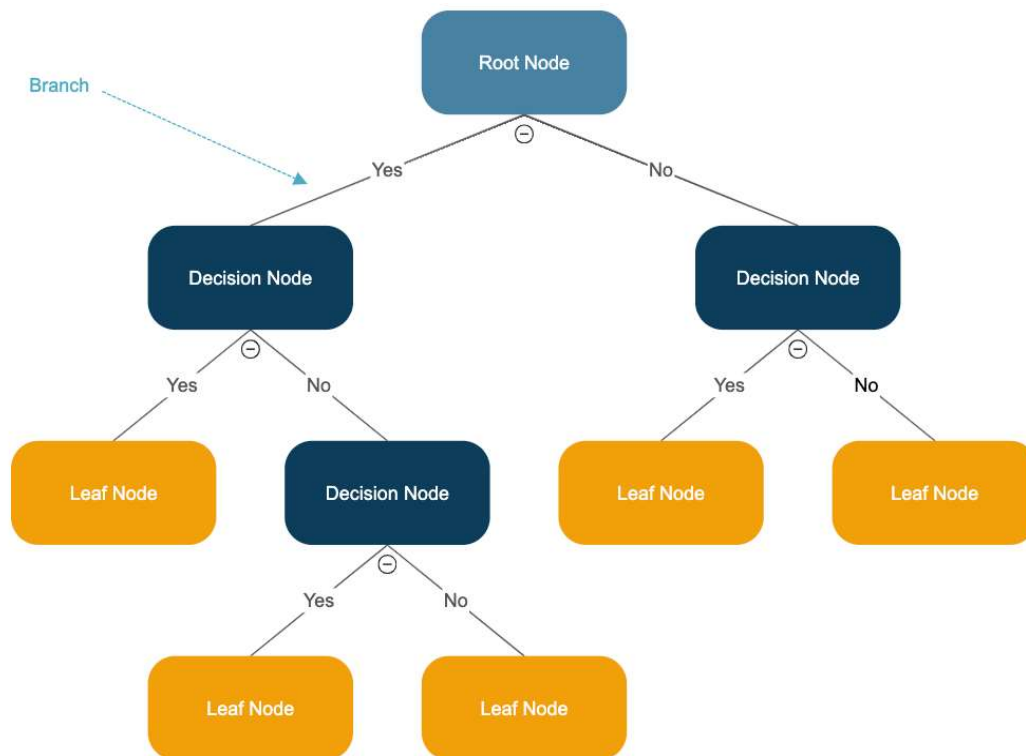
Platform used: Google Colab

Theory:

Decision Tree

A Decision Tree helps us to make decisions by mapping out different choices and their possible outcomes. It's used in machine learning for tasks like classification and prediction.

A Decision Tree helps us make decisions by showing different options and how they are related. It has a tree-like structure that starts with one main question called the root node which represents the entire dataset. From there, the tree branches out into different possibilities based on features in the data.



Root Node: Starting point representing the whole dataset.

Branches: Lines connecting nodes showing the flow from one decision to another.

Internal Nodes: Points where decisions are made based on data features.

Leaf Nodes: End points of the tree where the final decision or prediction is made.

There are mainly two types of Decision Trees based on the target variable:

Classification Trees: Used for predicting categorical outcomes like spam or not spam. These trees split the data based on features to classify data into predefined categories.

Regression Trees: Used for predicting continuous outcomes like predicting house prices. Instead of assigning categories, it provides numerical predictions based on the input features.

How Decision Trees Work?

1. Start with the Root Node: It begins with a main question at the root node which is derived from the dataset's features.

2. Ask Yes/No Questions: From the root, the tree asks a series of yes/no questions to split the data into subsets based on specific attributes.

3. Branching Based on Answers: Each question leads to different branches:

If the answer is yes, the tree follows one path.

If the answer is no, the tree follows another path.

4. Continue Splitting: This branching continues through further decisions helps in reducing the data down step-by-step.

5. Reach the Leaf Node: The process ends when there are no more useful questions to ask leading to the leaf node where the final decision or prediction is made.

Conclusion: Thus, we successfully applied the Logistic Regression for binary classification problems using machine learning, and assessed the model performance through appropriate evaluation metrics.