

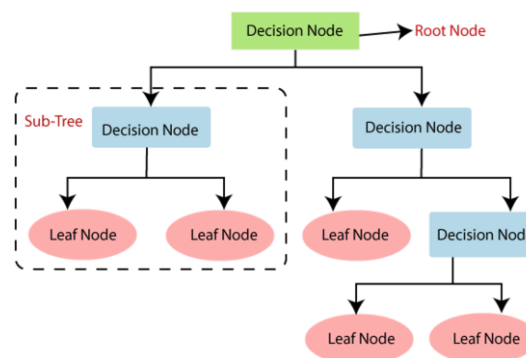
Decision Tree – 1 Page Summary

K V Mithilesh

AP23110011300

1. Introduction

A **Decision Tree** is a supervised ML algorithm that predicts an output by repeatedly splitting data based on the feature that provides the best separation. It is simple, interpretable, and works well for beginner-level classification tasks.



- It is like answering to Yes/No Questions
- Decision tree has root Node, leaf Nodes and internal Nodes

2. Dataset Used

We use a small dataset to predict **Transportation** based on:

- Gender
- Car Ownership
- Travel Cost
- Income Level

Target classes: **Bus, Train, Car**

Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

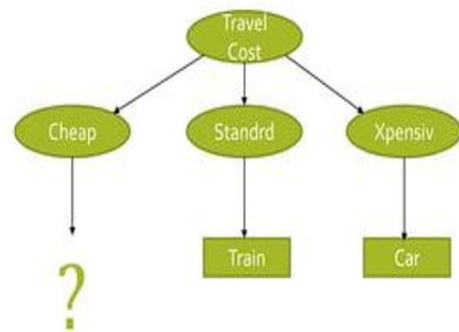
3. How the Tree Splits the Data (ID3 Intuition)

The algorithm chooses the feature with the **highest Information Gain** by calculating **Entropy**.

For this dataset:

Best first split: Travel Cost

- **Expensive** → Car
- **Standard** → Train
- **Cheap** → split again using Gender
 - Gender = Male → Bus
 - Gender = Female → Further Split

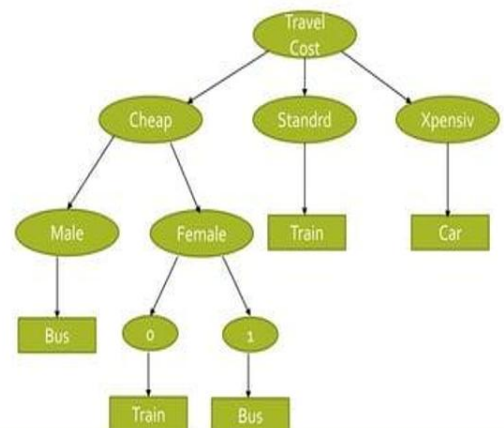


And do the splits until you get pure node.

4. Final Tree (Simplified)

Root: Travel Cost

- **Expensive** → Car
- **Standard** → Train
- **Cheap** →
 - Gender = Male → Bus
 - Gender = Female →
 - Car Ownership = 0 → Train
 - Car Ownership = 1 → Bus



5. Limitation & Fix

Limitation: Trees easily **overfit** small datasets.

Fix: Use **pruning**, limit depth, or switch to ensembles like **Random Forest**.

6. References

- [Ammar Mohammed, Decision Trees, SlideShare presentation.](#)
- [AlmaBetter, "Decision Tree in Machine Learning", Bytes Tutorial.](#)
- **Tom M. Mitchell**, *Machine Learning*, McGraw-Hill - Chapter on Decision Trees.