Decision Tree Exercise

1. Entropy of the Dataset

There are 10 records:

- Yes = 5
- No = 5

Entropy = 1.0

2. Information Gain (Short Summary)

Attribute	Information Gain
Weather	0.049
Road	0.049
Traffic	0.039
Engine Problem	0.396

Engine Problem has the highest Information Gain, so it is the root.

3. Gini Impurity

- Full dataset Gini = 0.5
- Engine = Yes → Gini = 0
- Engine = No → Gini = 0.408
- Weighted Gini = 0.286

Engine Problem is also the best according to Gini.

4. Final Decision Tree

Engine Problem?

/ \

Yes No

Yes (check Traffic)

/ \

High Other

Yes No

Interpretation:

- If Engine Problem = Yes → Accident = Yes
- If Engine Problem = No and Traffic = High → Accident = Yes
- Otherwise → Accident = No

Reflection

Based on the calculations, Engine Problem is the best attribute to split on because it has the highest Information Gain and the lowest Gini value. The decision tree shows that an engine problem strongly leads to an accident, and when there is no engine problem, high traffic increases the chance of an accident.