Week 8

I used online calculators to calculate entropy, entropy average, weighted entropy average, Information Gain, Gini Impurity, Gini Impurity average, and weighted Gini Impurity average.

1. Calculate Entropy and Information Gain:

Calculate Entropy for Entire Dataset:

E(D) = 1

Calculate Entropy for Each Attribute:

Weather Condition

E(Rain) = 0.92

E(Snow) = 0.92

E(Clear) = 1

Road Condition

E(Bad) = 0.81

E(Average) = 1

E(Good) = 0.81

Traffic Condition

E(High) = 0.82

E(Normal) = 0.92

E(Light) = 0.92

Engine Problem

E(Yes) = 0.81

E(No) = 0.92

Calculate the Weighted Entropy Average for Each Attribute:

Weather Condition = 0.95

Road Condition = 0.85

Traffic Condition = 0.88

Engine Problem Condition = 0.88

Calculate the Information Gain for Each Attribute:

Weather Condition = 0.05

Road Condition = 0.15

Traffic Condition = 0.12

Engine Problem = 0.12

Choose Best Attribute as Root Node:

Road Condition has highest information gain value.

Road Condition is root node.

2. Calculate Gini Impurity:

Calculate Gini Impurity for Entire Dataset:

Gini(D) = 0.5

Calculate Gini Impurity for Each Attribute:

Weather Condition

Gini(Rain) = 0.44

Gini(Snow) = 0.44

Gini(Clear) = 5

Road Condition

Gini(Bad) = 0.38

Gini(Average) = 0.5

Gini(Good) = 0.38

Traffic Condition

Gini(High) = 0.38

Gini(Normal) = 0.44

Gini(Light) = 0.44

Engine Problem

Gini(Yes) = 0.5

Gini(No) = 0.5

Calculate the Weighted Gini Impurity Average for Each Attribute:

Weather Condition = 0.47

Road Condition = 0.4

Traffic Condition = 0.42

Engine Problem Condition = 0.5

Choose Best Attribute as Root Node:

Road Condition has lowest Gini impurity value.

Road Condition is root node.

3. Construct the Decision Tree:

Determine first layer:

Root node = Road Condition

Branch 1 = Bad

Branch 2 = Average

Branch 3 = Good

Determine Branch 1 with Information Gain:

Weather Condition = 0.31

Traffic Condition = 0

Engine Problem = 0.31

Use Weather or Engine

Intermediate node = Weather Condition

Branch 1.1 = Rain

Branch 1.2 = Snow

Branch 1.3 = Clear

Determine Branch 2 with Information Gain:

Weather Condition = 0

Traffic Condition = 0

Engine Problem = 1

Intermediate node = Engine Problem

Branch 2.1 = Yes

Branch 2.2 = No

Determine Branch 3 with Information Gain:

Weather Condition = 0.31

Traffic Condition = 0

Engine Problem = 0.81

Intermediate node = Engine Problem

Branch 3.1 = No

Branch 3.2 = Yes

Final Decision Tree: