Name : Salman Almaskati

Income Level	Transportation Mode (Classes)
Low	Bus
High	Car
Low	Train
Medium	Bus
High	Car
Medium	Train
Medium	Bus
High	Car
Medium	Train
High	Bus
High	Car
Medium	Train

Calculate Entropy for the Dataset. Measure the information gain for the Income Level. Assume there are three other columns in dataset, explain how will you construct a decision tree.

Solution:

Entropy(Dataset) :

3 classes: Bus (4/12), Car 4/12, Train 4/12

Entropy = - 4/12 log2 (4/12) - 4/12 log2 (4/12) - 4/12 log2 (4/12) = **1.58496**

Information Gain (Income Level) :
3 classes: Low (2/12) , Medium (5/12), High (5/12)

Entropy(Income) = - (2/12) log2 (2/12) - (5/12) log2 (5/12) - (5/12) log2 (5/12) = **1.483**

Entropy(low) =- (1/2) log2 (1/2) - (1/2) log2 (1/2) - (0/5) log2 (0/5) = 0

Entropy(Medium) = - (2/5) log2 (2/5) - (3/5) log2 (3/5) - (0/5) log2 (0/5) = 0

Entropy(High) = - (4/5) log2 (4/5) - (1/5) log2 (1/5) - (0/5) log2 (0/5) = 0

Information gain = 1.58496 – (2/12)*0 – (5/12)*0 – (5/12)*0 = **1.58496**

To construct the decision tree you must first select the attribute for the root node, and create branches for possible attribute values, then split the instances into subsets, and repeat the process until all instances have the same class

Name : Salman Almaskati

Type of data	Level of measurement	Examples
Categorical	**Nominal** (no inherent order in categories)	Eye colour, ethnicity, diagnosis
	Ordinal (categories have inherent order)	Job grade, age groups
	Binary (2 categories – special case of above)	Gender
Quantitative (Interval/Ratio) (NB units of measurement used)	Discrete (usually whole numbers)	Size of household **(ratio)**
	Continuous (can, in theory, take any value in a range, although necessarily recorded to a predetermined degree of precision)	Temperature °C/°F (no absolute zero) **(interval)** Height, age **(ratio)**

Write down different data types provided in the table. Provide three examples for each of them (different from the examples provided in the table).

Categorical:

Nominal: Occupation, Race, Religion

Ordinal: Satisfaction rating **(like, neutral, dislike),** Grades **(A, B, C, D, F),** Severity levels **(Mild, Moderate, Severe, Critical)**

Binary**:** Smoking **(Smoker/ non-smoker) ,** Medical test **(positive or negative),** Employment status **(Employed/Unemployed)**

Quantitative:

Discrete: Number of siblings, Number of pets, Number of students

Continuous: Distance, Weight, Speed