1 Datasets

ID	Age	CreditScore	Education	RiskLevel
1	35	720	16	Low
2	28	650	14	High
3	45	750	missing	Low
4	31	600	12	High
5	52	780	18	Low
6	29	630	14	High
7	42	710	16	Low
8	33	640	12	High

Table 1: Training Dataset (8 records)

ID	Age	CreditScore	Education
T1	37	705	16
T2	30	645	missing

Table 2: Test Dataset (2 records)

2 Question 1

Let us define A as the initial training dataset before the split, with 8 records: 4 in class $High$ and 4 in class Low. We can compute the entropy of A, denoted by $H(A)$, as:

$$H(A) = -\frac{4}{8}\log_2\frac{4}{8} - \frac{4}{8}\log_2\frac{4}{8} = 1$$

Splitting A on $CreditScore$ at 650, we obtain the following results:

- Left node ($CreditScore \leq 650$): IDs 2, 4, 6, 8 → All are $High$
- Right node ($CreditScore > 650$): IDs 1, 3, 5, 7 → All are Low

Without calculating the entropy of the dataset after the split, we can immediately obtain the information gain of the split as 1. This is because the dataset is completely pure after the split (all records with $CreditScore$ less than or equal to 650 have $HighRiskLevel$ and all records with $CreditScore$ greater than 650 have $LowRiskLevel$), therefore, the information gain equals the initial entropy before the split.

However, we can calculate the entropy of the dataset after the split as follows:

$$H(\text{CreditScore} \leq 650) = -\frac{4}{4}\log_2\frac{4}{4} = 0$$

$$H(\text{CreditScore} > 650) = -\frac{4}{4}\log_2\frac{4}{4} = 0$$

$$\text{InformationGain}(\text{CreditScore} = 650) = H(A) - \frac{4}{8}\cdot 0 - \frac{4}{8}\cdot 0 = 1 - 0 - 0 = 1$$

Since the information gain is maximized, we would choose this as a root node split.

3 Question 2

The main difference between regression trees and classification trees lies in the type of output they produce. Regression trees are used to predict continuous numerical values, whereas classification trees are designed to predict categorical classes. In regression trees, the leaf nodes represent the average value of the target variable within that node, minimizing prediction error. In contrast, the leaf nodes of classification trees represent the most probable class label. Additionally, regression trees typically use variance reduction as the splitting criterion, while classification trees use measures like information gain or entropy to determine the best splits. Table 3 provides a detailed comparison between regression trees and classification trees.

Aspect	**Regression Tree**	**Classification Tree**
Output Type	Continuous values	Categorical classes
Leaf Node Value	Average of target values	Most probable class
Splitting Criterion	Variance reduction	Information gain / Entropy
Purpose	Predict numerical outcomes	Classify into categories
Error Metric	Mean Squared Error (MSE)	Classification Error / Gini / Entropy

Table 3: Comparison between Regression Trees and Classification Trees

To compute the variance reduction when splitting A on Age at 35, we first compute the variance of A as follows:

$$Mean(A) = \frac{720 + 650 + 750 + 600 + 780 + 630 + 710 + 640}{8} = \frac{5480}{8} = 685$$

$$\begin{aligned} Var(A) &= \frac{1}{8}\Big[(720-685)^2 + (650-685)^2 + (750-685)^2 \\ &\quad + (600-685)^2 + (780-685)^2 + (630-685)^2 \\ &\quad + (710-685)^2 + (640-685)^2\Big] \\ &= \frac{1}{8}\Big[1225 + 1225 + 4225 + 7225 + 9025 + 3025 + 625 + 2025\Big] \\ &= \frac{38700}{8} = 4837.5 \end{aligned}$$

The variance of the dataset after splitting can be computed as follows:

$$Mean(Age \leq 35) = \frac{720 + 650 + 600 + 630 + 640}{5} = \frac{3240}{5} = 648$$

$$\begin{aligned} Var(Age \leq 35) &= \frac{1}{5}\Big[(720-648)^2 + (650-648)^2 + (600-648)^2 \\ &\quad + (630-648)^2 + (640-648)^2\Big] \\ &= \frac{1}{5}\Big[5184 + 4 + 2304 + 324 + 64\Big] \\ &= \frac{7880}{5} = 1576 \end{aligned}$$

$$Mean(Age > 35) = \frac{750 + 780 + 710}{3} = \frac{2240}{5} \approx 746.67$$

$$\begin{aligned}
Var(Age > 35) &= \frac{1}{3}\Big[(750 - 746.67)^2 + (780 - 746.67)^2 + (710 - 746.67)^2\Big] \\
&\approx \frac{1}{3}\Big[11.1 + 1111.1 + 1344.4\Big] \\
&= \frac{2466.6}{3} \\
&\approx 822.2
\end{aligned}$$

$$Var(Age = 35) = \frac{5}{8} \cdot 1576 + \frac{3}{8} \cdot 822.2 = 985 + 308.33 = 1293.33$$

Finally, we can compute the variance reduction when splitting A on Age at 35 as follows:

$$VarianceReduction(Age = 35) = 4837.5 - 1293.33 = 3544.17$$

4 Question 3

// To do

5 Question 4

// To do

6 Question 5

// To do

7 Question 6

// To do

8 Question 7

// To do

9 Question 8

// To do

10 Question 9

// To do

11 Question 10

// To do

12 Question 11

// To do

13 Question 12

// To do

14 Question 13

// To do

15 Question 14

// To do

16 Question 15

// To do

17 Question 16

// To do

18 Question 17

// To do

19 Question 18

// To do

20 Question 19

// To do

21 Question 20

// To do