ourdiu.com | Shared by SUPAN ROY

CT-2 CSE 431 Machine Learning (Sec-C)

Total Marks: 15 Time: 30 Minutes

1. A digital streaming company wants to predict whether a visitor will purchase a premium subscription after viewing advertisements. The dataset includes the following information:

ID	Age Group	Monthly Income	Ad Viewing Time (Minutes)	Purchase Premium
1	Young	High	12	Yes
2	Young	Low	5	No
3	Young	Low	8	No
4	Adult	High	15	Yes
5	Adult	Low	10	Yes
6	Adult	High	18	Yes

a) Calculate the entropy of the parent dataset.

b) Split the dataset using Age Group and calculate the entropy of each child node, weighted entropy, and information gain.

c) Consider the numerical attribute Ad Viewing Time. Use a threshold of 9 minutes to divide the dataset into two groups (≤ 9 minutes and > 9 minutes). Calculate the entropy of each child node, weighted entropy, and information gain.

d) *Compare the information gains of Age Group and Ad Viewing Time. Which attribute should be selected as the root node? Justify your answer.*