



## EAST WEST UNIVERSITY

### Department of Computer Science and Engineering

#### Assignment, Summer 2024 Semester

**Course:** Machine Learning, CSE475/ ICE478, Section - 2  
**Instructor:** Dr. Raihan Ul Islam, Associate Professor, Department of CSE  
**Full Marks:** 15  
**Time:** 10<sup>th</sup> September, 24

#### Notes:

- C2: Understand, use, and determine appropriate machine learning methods/algorithms suitable for different types of learning problems, i.e. know about their most important weaknesses and advantages.
- There are 3 (Three) questions, answer them. Mark of each question are mentioned at the right margin.

1. Explain the concept of entropy and information gain in the context of decision trees. [C2
2. From the following dataset predict the loan default status. Construct a decision tree using the ID3 algorithm. Show the calculations for entropy and information gain at each step. Marks: 2+10+3 = 15]

Age	Income	Education	Credit Score	Loan Default
Young	Low	High School	Fair	Yes
Young	Medium	Bachelor's	Good	No
Middle-aged	High	Master's	Excellent	No
Senior	Low	High School	Poor	Yes
Young	High	Bachelor's	Excellent	No
Middle-aged	Medium	Bachelor's	Fair	Yes
Senior	High	Master's	Good	No
Young	Low	High School	Poor	Yes
Middle-aged	High	Master's	Excellent	No
Senior	Medium	Bachelor's	Good	No
Young	Medium	High School	Fair	Yes
Middle-aged	Low	Bachelor's	Poor	Yes
Senior	High	Master's	Excellent	No
Young	High	Master's	Good	No
Middle-aged	Medium	High School	Fair	No

3. Overfitting isn't just a theoretical concern; it has real-world consequences. Discuss the potential implications of overfitting in a decision tree model deployed in a critical application like medical diagnosis or financial forecasting. How might overfitting impact the model's reliability, and the decisions made based on its predictions?