Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science and Engineering
Mid Semester Examination, Spring-2024
Course Code: CSE321 Course Title: Data Mining and Machine Learning
Level: 3 Term: 2

Exam Duration: 1.5 Hours **Marks: 25**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1. *a)* Discuss any two categories of tasks usually performed in Data Mining and Machine Learning with example. **[3]** **CO1**

b) Explain the impact of "Curse of dimensionality" in Data Mining and Machine Learning tasks. **[2]**

2. Consider the following data set of students and their scores in three subjects: **CO2**

Name	Math	Physics	Chemistry
Rahim	85	90	95
Karim	75	80	85
Julekha	65	70	75
Sokhina	55	60	65
Sakib	63	55	70
Mysha	70	78	83
Zorina	90	87	82
Arman	63	58	52

a) Apply simple linear regression to find the equation of the best-fit line that predicts the Physics score based on the Math score. Show your calculations and the final equation. **[3]**

b) Use the equation of the best-fit line calculated in 2a. to predict the Physics score of a student who scored 70 in Math. Show your calculation and the final answer. **[2]**

3. Consider the following classification task of whether a person is fit or not based on their age and weight. **[5]** **CO2**

Age	Weight	Class
30	80	Fit
40	90	Not fit
35	60	Fit
45	70	Not fit
38	73	Fit
30	80	Fit
20	90	Not fit

Use any classification algorithm to classify whether a person who is 25 years old and weighs 70 kg is fit or not fit.

4. Consider the following dataset of transportation information of employees of a company.

Gender	Travel Cost	Income Level	Transportation Mode
Male	Cheap	Low	Bus
Male	Cheap	Medium	Bus
Female	Cheap	Medium	Train
Female	Cheap	Low	Bus
Male	Cheap	Medium	Bus
Male	Standard	Medium	Train
Female	Standard	Medium	Train
Female	Expensive	High	Car
Male	Expensive	Medium	Car
Female	Expensive	High	Car
Male	Standard	Medium	?
Female	Cheap	Medium	?

CO3

a) Build a decision tree model from the given dataset based on entropy. You should only consider the labeled data for building the tree. **[8]**

b) Predict the class of the unlabeled data in the dataset using the model built in Q4(a). **[2]**

$$E = -\frac{1}{total}$$

$$\frac{A}{A+B+C}$$