

International Islamic University Chittagong

Department of Computer Science and Engineering

Midterm, Spring-2023

Course Code: CSE-4877 Course Title: Machine Learning and Data Mining

Total marks: 30 Time: 1.5 hours

[Answer the following questions. Figures in the right-hand margin indicate full marks.]

- 1.a) Discuss (shortly) whether each of the following activities is a data-mining task or not. 2 CLO2
- i. Predicting the outcomes of tossing a (fair) pair of dice.
 - ii. Predicting the future stock price of a company using historical records.
 - iii. Recommending videos for the YouTube users.
 - iv. Monitoring seismic waves for earthquake activities.
- b) Suppose you have the following data set: 5 CLO2

	A1	A2
X1	1.4	1.6
X2	2.1	1.9
X3	1.7	1.8
X4	1.3	1.6
X5	1.6	1.0

Given a new point, $X = (1.3, 1.5)$ as a query, rank the database points based on similarity with the query using Euclidean distance, Manhattan distance, supremum distance and cosine similarity.

- c) Consider the following data: 12, 14, 15, 15, 18, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40, 46, 47, 53, and 72. 3 CLO3
- i. What is the mode of the data? Comment on the data's modality.
 - ii. Give the five-number summary of the data.
 - iii. Show a boxplot of the data.
2. Take a look at the data table below, which contains mixed-type variables. Display the dissimilarity matrix between the mixed-type variables. Remember to set $P=1$ and $N=0$ when dealing with binary variables. Additionally, a binary variable is thought to be asymmetric. 10 CLO3

ID	Attribute-1 (categorical)	Attribute-2 (binary)	Attribute-3 (ratio-scaled)
1	Red	P	11
2	Blue	P	22
3	Green	N	64
4	Red	P	110

- 3.a) Apply Apriori algorithm with a minimum support threshold of 2 to the following transaction database: 7 CLO3

TID	Items
T1	I1, I2, I5
T2	I2, I4
T3	I2, I3
T4	I1, I2, I4
T5	I1, I3

- b) i. List all frequent itemset returned by Apriori algorithm
By considering Q.3.(a), generate all interesting rules those minimum confidence is 60%.

3 CLO2

OR

- 3.a) Consider the following dataset where Quantity Demand depends on the Price and Income of the product. Use linear regression to predict the value of Quantity Demand when Price and Income are respectively 6 and 600.

6 CLO2

Price (X1)	Income (X2)	Quantity Demand (Y)
5	1000	100
7	600	75
6	1200	80
6	500	70
8	300	50
7	400	65
5	1300	90
4	1100	100
3	1300	110
9	300	60
6	600	?

- b) How the error is calculated in a Linear Regression model? Explain with an example.

4 CLO3