ourdiu.com | Shared by Mehedi Hasan Jeon 2.4222E+14

Daffodil International University

Faculty of Science & Information Technology

Department of Computer Science & Engineering

Mid Examination, Summer 2025

Course Code: CSE325, Course Title: Data Mining and Machine Learning

Level: 3 Term: 3 Batch: 63

Time: 01:30 Hrs **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.]

In the pulse of Bangladesh, between Dhaka's chaos and Rajshahi's calm, five lives moved forward, marked by time and savings. Rajib, 1.1 years into his tech venture in Rajshahi, had saved 39.3. Every thousand-taka note told of long nights and stubborn hope. In Dhaka, Sharmeen, a freelance designer, had been working for 1.3 years with 46.2 set aside. Her hustle never stopped; every client won was a small victory. Back in Rajshahi, Nasim, a teacher and father, had saved 37.7 in 1.5 years. Quiet and determined, his dreams stretched beyond borders for his daughter's future. Anika, two years into her career in Dhaka, had 43.5 saved. By day, she balanced spreadsheets; by night, she nurtured a secret dream of starting a startup. Then came Zayaan, 3.2 years into his journey. No one knew where he came from or what he had, but his eyes spoke of lessons hard-earned and unfinished plans. Five stories. Five numbers. One shared struggle: to turn thousands into a life worth living.

Now, based on the story above, answer the following questions 1 and 2.

(Note: All financial values are given in thousands; for calculations, you may treat them as it is.)

1	*a)*	**Organize the** table based on the given data and **identify** any missing values. If some of these missing values are Missing Not At Random (MNAR), what kind of preprocessing techniques can be applied to handle them? Please **explain** your reasoning.	2	**CO1**
	b)	**Explain** how measuring distance or similarity between data points can assist in handling missing values in a dataset.	3	
2.	*a)*	**Select** appropriate features and **estimate** the missing financial value using a method that models a continuous trend over time. After that, **use** an appropriate classifier to **find** the missing location	7	**CO2**
	b)	**Justify** the rationale behind your model selection in part 2 a). How would you **evaluate** their performance using suitable metrics? **Explain** with examples for each.	3	

ourdiu.com | Shared by Mehedi Hasan Jeon 2.4222E+14

Employee	Years in Company	Monthly Sales (units)
A	2.0	150
B	3.5	200
C	1.2	120
D	4.0	210
E	2.5	170

3.	*a)*	**Find** the first and second (pc1 and pc2) principal components for the above dataset.	7	**CO2**
	b)	**Compute** the Mahalanobis distance, and **explain** why the Mahalanobis distance might be **preferred** over the Euclidean distance in analyzing the above dataset. Employee A: (2.0 years, 150 sales) Employee C: (1.2 years, 120 sales)	3	