

THE OPEN UNIVERSITY OF SRI LANKA
 BSc Hons (IT) DEGREE PROGRAMME: LEVEL 06
 FINAL EXAMINATION: SEMESTER 2: 2024/2025
 COU6306: DATA MINING
 DURATION: TWO HOURS (2 HOURS)



Date: 16.11.2025

Time: 1.30 pm – 3.30 pm

Answer FOUR Questions ONLY.

Q1.

- A. Explain the significance of data mining in today's data-driven world. [03 Marks]
- B. What are the major data types involved in data mining? Briefly explain each. [06 Marks]
- C. Describe any three (03) data mining functionalities. [06 Marks]
- D. Define each of the following terms. [10 Marks]
- (i) Machine Learning
 - (ii) Information Retrieval
 - (iii) Statistical Hypothesis Test
 - (iv) Semi-supervised Learning
 - (v) Active Learning

Q2.

- A. Briefly explain two (02) popular and successful applications of data mining. [04 Marks]
- B. Define the term "correlation" and list the three (03) types of correlation. [06 Marks]
- C. Explain three (03) social implications of data mining. [06 Marks]
- D. What is principal component analysis (PCA)? List the six (06) steps involved in PCA and describe its uses. [09 Marks]

Q3.

A. State the formulas used to calculate standard deviation and variance.

[04 Marks]

B. Draw the box plot and clearly label each part.

[05 Marks]

C. List and explain the measures of central tendency.

[06 Marks]

D. You are given the following dataset:

Dataset = {5, 8, 12, 6, 10, 9, 7, 14, 11, 13, 9, 15}

Calculate the Following:

- (i) Arithmetic Mean
- (ii) Harmonic Mean
- (iii) Median
- (iv) Mode
- (v) Range

[10 Marks]

Q4.

A. State three (03) methods to handle missing values in data mining.

[03 Marks]

B. List four (04) data visualization techniques.

[04 Marks]

C. Identify and briefly describe the four (04) major tasks in data preprocessing.

[08 Marks]

D. Provide the equations for each of the following:

- (i) Dissimilarity between nominal attributes
- (ii) Chebyshev distance
- (iii) Weighted Euclidean distance
- (iv) Similarity between nominal attributes
- (v) Dissimilarity for attributes of mixed types

[10 Marks]

Q5.

-
- A. Define the word "data integration".
[03 Marks]
- B. What is the Entity Identification Problem? Give a real-world example.
[06 Marks]
- C. List and explain the three (03) data reduction techniques.
[06 Marks]
- D. Identify and define the four (04) approaches to discretization by binning, providing an example for each.
[10 Marks]

Q6.

- A. List the three (03) main steps of the apriori algorithm.
[03 Marks]
- B. Explain the three-tier architecture of the data warehouse.
[06 Marks]
- C. State three (03) differences between Operational database systems (OLTP) and Data warehouses (OLAP).
[06 Marks]
- D. List the four (04) steps of the data warehouse design process and briefly describe each.
[10 Marks]

*** *All Rights Reserved* ***

