

PAN AFRICA CHRISTIAN UNIVERSITY

**BACHELOR OF BUSINESS INFORMATION TECHNOLOGY AND BACHELOR
OF SCIENCE IN INFORMATION TECHNOLOGY**

END OF SEMESTER EXAMINATION

DEPARTMENT: COMPUTING AND INFORMATION TECHNOLOGY

COURSE CODE: BSIT307

COURSE TITLE: INTRODUCTION TO DATAMINING

CAMPUS: ROYSAMBU

EXAM DATE: THURSDAY, ,2024

TIME: 8:00 AM - 11:00AM

INSTRUCTIONS

- This exam script has **TWO (2)** sections.
- Read all questions carefully before attempting.
- Answer All questions in Section **A** and any other Three questions in Section **B**.
- Write only your **student number** on the answer booklet provided.

SECTION A: (Answer ALL questions in this section)

Question 1:

- a) Data mining is a valuable tool in data warehousing as it allows organizations to discover hidden insights and patterns in large amounts of data to inform decision -making and improve business outcomes. Using examples explain three types of datamining techniques

[6 Marks]

- b) "A data warehouse is typically designed to be read-only database". Using an example explain the statement as used in data warehousing

[2

Marks]

- c) Discuss the two major tasks in data preprocessing as used in data mining pipeline

[2

Marks]

SECTION B: (Answer any THREE (3) questions in this section)

Question 2:

- a) ABX Corporation is a multinational company that operates in several countries around the world. The company's management team wants to gain deeper insights into its sales operations to better understand customer behavior and make more informed business decisions. To achieve this goal ABX Corporation decides to implement a data warehouse. Discuss the THREE key characteristics of a data warehouse that the company should consider to achieve its objective.

[6

Marks]

- b) Distinguish between "Feature selection" and "Model selection" as used in data mining

[4

Marks]

Question 3:

- a) A telecommunication company, XYZ Telecom, is experiencing a high rate of customer churn. They want to implement a data mining solution to predict which customers are likely to churn, so they can take proactive measures to retain them. XYZ company has collected a vast amount of customer data, including demographics, service usage, billing information and customer complaints. They also have a historical churn data set that includes customers who have churned in the past:

Required:

- i. Outline the data preprocessing steps you would undertake to prepare the data for churn prediction
[5 Mark]
- ii. Explain two techniques that you would use to handle missing values **[2 Mark]**
 - b) Analyze the impact of data quality issues on the effectiveness of a data warehouse **[2 Mark]**
 - c) Explain the purpose of data warehousing in an organization **[1 Mark]**

Question 4:

- a) Online Analytical Processing (OLAP) operations are designed to facilitate multidimensional analysis of data in a data warehouse. Discuss five OLAP operations as used in data warehousing **[5 Marks]**
- b) Describe how decision tree algorithm works as used in knowledge discovery **[5 Marks]**

Question 5:

- a) Data warehouses use dimensional model design. Describe two data warehouse schema as used in logical design in data warehouse environment illustrating with diagrams **[6 Marks]**
- b) Differentiate between "data marts" and "data warehouses" as used in data warehousing and data mining **[4 Marks]**

Question 6:

- a) Discuss any two applications that use data mining **[2 Marks]**
- b) Classify the data mining systems based on the data mining functionalities **[6 Marks]**
- c) "Data is dirty and need to be cleaned before being loaded into the data warehouse" Explain the statement as used in data warehousing **[2 Marks]**

**PAN AFRICA CHRISTIAN UNIVERSITY
EXAMINATIONS PAPER**



EOF