INTERNATIONAL UNIVERSITY OF EAST AFRICA

'Learning to Succeed'

FACULTY OF SCIENCE AND TECHNOLOGY

END OF SEMESTER EXAMINATIONS - APRIL 2025

PROGRAMME: MIT

YEAR/SEM: YEAR 1/SEMESTER 2

COURSE CODE: MIT721

NAME: BUSINESS INTELLIGENCE & DATA MINING

DATE: 2025-04-22

TIME: 2:00-5:00PM

INSTRUCTIONS TO CANDIDATES:

1. Read the instructions very carefully

2. The time allowed for this examination is STRICTLY three hours

3. Read each question carefully before you attempt and allocate your time equally between all the Sections

4. Write clearly and legibly. Illegible handwriting cannot be marked

5. Number the questions you have attempted

6. Use of appropriate workplace examples to illustrate your answers will earn you bonus marks

7. Any examination malpractice detected will lead to automatic disqualification.

DO NOT WRITE ANYTHING ON THE QUESTION PAPER

Section A Attempt all questions

Question 1:

A software development company, *Kampala Tech Solutions*, is building an AI-powered analytics tool for businesses. The tool aims to help businesses make better decisions by analyzing large datasets collected from their operations, customers, and market trends. The tool is designed to handle different types of tasks like predicting sales, classifying customer feedback, clustering customer segments, and more. The development team at *Kampala Tech* is currently discussing various machine-learning approaches to implement these features efficiently. They need to decide whether to use supervised or unsupervised learning and choose appropriate algorithms based on the specific problems they are solving.

Based on the case study above, answer the following questions:

(a) The team is debating whether to use supervised or unsupervised learning for different modules in the tool. Explain the major difference between supervised and unsupervised learning, and suggest one suitable machine learning application within *Kampala Tech's* tool for each approach. *(10)*

(b) One of the modules in the tool is designed to predict the future revenue of clients, while another is designed to categorize incoming customer emails into different departments (e.g., sales, support, complaints). Explain how a regression problem is different from a classification problem, using these two modules as examples.

(10 marks)

(c) Identify and explain any one software engineering problem that *Kampala Tech* might encounter, which would require:

    (i) A regression algorithm (8 marks)
    (ii) A classification algorithm (8 marks)

(d) Recommend one specific machine learning algorithm that the team could implement for:
    (i) Regression (2 marks)
    (ii) Classification *(2 marks)*

Section B Attempt a maximum of 3 questions.

Question 1:

(a) What is K in the KNN (k Nearest neighbor) algorithm? (4 marks)
(b) Is KNN a supervised or unsupervised learning algorithm? Explain your answer (4 marks)
(c) Explain the steps for performing classification with the KNN algorithm (8 marks)
(d) How can KNN be used for a regression task? (4 marks)

Question 2:

(a) Explain the importance of data preprocessing in data mining. (8 marks)
(b) Describe any two data techniques for scaling data (8 marks)
(c) How is missing data handled in data mining (4 marks)?

Question 3:

((a) Explain the importance of data preprocessing in data mining. (10 marks)
(b (b) State whether the following real-world problems require **classification or regression or clustering**. (Write either **classification** or **regression,** or **clustering** against each problem). (2 mark each)

(i) Estimating the price of a house based on its features like size and location.

(ii) Predicting the annual salary of an employee based on experience and education.

(iii) Diagnosing whether a tumor is benign or malignant based on medical images.

(iv) Predicting the number of units a store will sell next month based on past sales data.

(v) Segmenting customers into different groups based on purchasing behavior.entifying different speech patterns in voice recognition.

Question 4:

Explain the following Data Warehouse Back-End Tools and Utilities. (4 marks each)

(a (a) Data extraction:

(b (b) Data cleaning:

(c (c) Data transformation:

(d (d) Load:

(e (e) Refresh

Question 5:

(a)Â With examples, explain how categorical data is different from continuous dataÂ (10 marks)

(b) Explain how categorical data is handled during data preprocessing |(10 marks)

Question 6:

Consider the data for predicting car prices given in Fgure1 below.Â Price is the target variable.

Brand	Price (US$)	Age (years)	Distance (KM)	Weight (kgs)
BMW	13500	23	46986	1165
Toyota	13750	23	72937	1165
Honda	13950	24	41711	1165
BMW	14950	26	48000	1165
BMW	13750	30	38500	1170
Toyota	12950	32	61000	1170
Honda	16900	27	94612	1245
BMW	18600	30	75889	1245
Honda	21500	27	19700	1185
BMW	12950	23	71138	1185

Figure 1: Predicting Car Prices

(a)Â Â Â Which data miningÂ task is problem, regression or classification ? Why?Â (4 marks)

(b)Â Â Least the features in the dataset for solving the problem.Â (3 marks)

(c)Â How many data samples (instances) is this dataset?Â (2 marks)

(d)Â

(i) Assume we want to predict the Car Brand, using the rest of other attributes. What machine problem will best fit this task?Â Explain why.Â Â (4 marks)

(ii) List the labels in this case (3 marks)

(e) Explain any two data preprocessing tasks that can be performed on this dataset. (4 marks)