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:
TIME:

INSTRUCTIONS TO CANDIDATES:

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)*

Question 2:

. Kansanga Mart Ltd. is a fast-growing retail chain that operates both physical stores and an online platform. Over the past few years, the company has been collecting large volumes of data from sales transactions, customer feedback, inventory systems, and market trends. However, the management team has been facing challenges in making strategic decisions due to the lack of proper data analysis and reporting tools.

To gain a competitive edge and improve decision-making, *Kansanga Mart Ltd.* decides to invest in a **Business Intelligence (BI)** system. They plan to use BI to transform raw data into meaningful insights that can guide their marketing strategies, inventory management, and customer satisfaction initiatives.

The company’s IT department has been tasked with understanding the concept of Business Intelligence and designing a suitable BI framework for implementation.

(a) Explain the concept of Business Intelligence and how it could benefit *Kansanga Mart Ltd.* in improving its decision-making process.

(4 marks)

(b) With the aid of a well-labeled diagram, illustrate a typical Business Intelligence framework that could be implemented by *Kansanga Mart Ltd.*

(6 marks)

(c) Identify and explain any five major components of the Business Intelligence framework illustrated in part (b), highlighting how each component would assist *Kansanga Mart Ltd.* in achieving its goals.

(10 marks)

SECTION B

Attempt a maximum of 3 questions.

Question 1:

i. (a) Explain the following terms in data mining, (4 marks each)

i. Regression,

ii. Classification,

iii. Clustering problem.

(b) List two algorithms in each of the tasks in (a) above (2 marks each)

(c) Why is data mining important? (2 marks)

Question 2:

Consider the data for predicting car prices given in Fgure1 below. 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) Which data mining task is problem, regression or classification ? Why? (4 marks)

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

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

(d)

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

Question 3:

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

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

Question 4:

(a) With the aid of diagram examples, Explain the concept of data splitting. (10 marks)

(b) write short notes about the following

(i) Test data

(ii) Train data

Question 5:

(a) What is a Data Warehouse? (2 marks)

(b) Explain its characteristics. (8 marks)

(c) Compare Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP). (10 marks)

Question 6:

((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 7:

Explain the followingData 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 8:

(a) Explain the following Data Warehouse models. (5 marks each)

(i) Enterprise warehouse

(ii) Data Mart

(iii) Virtual warehouse

(b) What is Online Analytical Processing (OLAP)? Explain its significance in BI. (5 marks)

Question 9:

(a) Bias and variance are common concerns when working on data mining tasks. Explain the difference between variance and bias (4 marks)

(b) Explain any four ways to minimize overfitting (8 marks)

(c) Detail four ways to overcome under-fitting (8 marks)

Question 10:

Consider the following training data for predicting students' results

Score1	29	22	10	31	17	33	32	20
Score2	43	29	47	55	18	54	40	41
Result	Pass	Fail	Fail	Pass	Fail	Pass	Pass	Pass

Figure 2. Predicting students' results

(i) What machine learning problem best fits this data? Classification or regression? Explain your answer (4 marks)

(ii) How many features is the dataset? List them. (4 marks)

(iii) How many data samples/instances is this data? (4 marks)

(iv) What is the target variable in this ? (4 marks)

(v) Mention the labels in the dataset (4 marks)

Question 11:

(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 12:

(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)