INTERNATIONAL UNIVERSITY OF EAST AFRICA
'Learning to Succeed'

FACULTY OF SCIENCE AND TECHNOLOGY

END OF SEMESTER EXAMINATIONS - AUGUST 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:

KabalagalaKing Hardware Ltd., a construction materials and hardware supply company based in Kampala, is integrating an AI-powered analytics system to enhance its business decision-making processes. This system is intended to help the company analyze large volumes of data related to product sales, customer orders, inventory levels, and market trends.

The AI tool is expected to perform various tasks, such as predicting future demand for building materials, categorising customer inquiries, and segmenting clients based on their purchasing behaviour. As the system is being developed, KabalagalaKing's data analytics team is evaluating which machine learning approaches are most suitable for each task and which algorithms should be used accordingly.

(a) Explain the fundamental difference between supervised and unsupervised learning. Provide one example of how each approach could be applied within the business intelligence system at KabalagalaKing Hardware Ltd. (8 marks)

(b) One module of the AI system is aimed at forecasting future revenue based on past sales trends, while another is developed to sort customer emails into categories such as quotations, complaints, deliveries, or technical support. Use these examples to distinguish between a regression problem and a classification problem. Clarify the module to be handled as a classification task and a regression one. (8 marks)

(c) Identify and explain one practical business problem that KabalagalaKing Hardware Ltd. might face which would require:

(i) A regression algorithm (5 marks),

(ii) A classification algorithm (5 marks)

(iii) A Classification algorithm (5 marks)

(d) Suggest and briefly explain one specific data mining algorithm that could be implemented to solve:

(i) Regression problem (3 marks),

(ii) For solving a classification problem (3 marks)

(iii) Clustering problem (3 marks) for KabalagalaKing Hardware Ltd.

Question 2:

NsambyaMic Ltd. is a microfinance institution based in Uganda, offering small loans, savings accounts, and financial advisory services to low-income individuals and small businesses. To enhance decision-making and improve customer service, NsambyaMic is investing in a Business Intelligence (BI) system with integrated data mining capabilities.

The BI system will analyze customer data such as loan repayment histories, savings patterns, demographic details, and transaction behavior. The goal is to help the company predict loan defaults, categorize loan applications, identify groups of similar clients, and monitor overall financial performance.

NsambyaMic's data team is exploring the use of various data mining techniques such as regression, classification, and clustering to support different decision-making processes.

(a) Briefly explain the term Business Intelligence (BI) and describe two ways it can support decision-making at NsambyaMic Ltd. (8 marks)

(b) Consider the following two tasks:

(i) Predicting whether a new loan applicant is likely to default or repay the loan;

(ii) Predicting the amount a customer is likely to save each month.

For each task, Identify whether it is a classification or regression problem. Justify your answer. Explain the importance of each task towards increasing the business revenue (4 marks @)

(c) NsambyaMic wants to group its customers into categories based on their loan size, repayment speed, savings frequency, and income level in order to tailor financial products more effectively. Which data mining technique is most suitable for this task? Explain its purpose and benefits in this microfinance context. (3 marks)

Provide one other different realistic example of how NsambyaMic could use each of the following techniques: (i) Regression, (ii) Classification, (iii) Clustering. In each case, briefly describe the data used and the business goal. (4 marks each)

Suggest one appropriate machine learning algorithm for each of the following tasks and briefly explain why it is suitable: (i) Predicting the amount a client will save in the next three months, (ii) Classifying whether a loan application is low-risk or high-risk, (iii) Segmenting clients into behavioral groups based on their financial activity. (3 marks each)

SECTION B

Attempt a maximum of 3 questions.

Question 1:

(a) Describe the concept of Business Intelligence. (4 marks)

(b) Using a diagram, illustrate the Business Intelligence framework. (10 marks)

(c) Discuss any three key components of the Business Intelligence framework you presented in part. (6 marks)

Question 2:

NsambyaMic Ltd. is a microfinance institution based in Uganda, offering small loans, savings accounts, and financial advisory services to low-income individuals and small businesses. To enhance decision-making and improve customer service, NsambyaMic is investing in a Business Intelligence (BI) system with integrated data mining capabilities.

The BI system will analyze customer data such as loan repayment histories, savings patterns, demographic details, and transaction behavior. The goal is to help the company predict loan defaults, categorize loan applications, identify groups of similar clients, and monitor overall financial performance.

NsambyaMic's data team is exploring the use of various data mining techniques such as regression, classification, and clustering to support different decision-making processes.

(a) Briefly explain the term Business Intelligence (BI) and describe two ways it can support decision-making at NsambyaMic Ltd. (8 marks)

(b) Consider the following two tasks:

(i) Predicting whether a new loan applicant is likely to default or repay the loan;

(ii) Predicting the amount a customer is likely to save each month.

For each task, Identify whether it is a classification or regression problem. Justify your answer. Explain the importance of each task towards increasing the business revenue (4 marks @)

(c) NsambyaMic wants to group its customers into categories based on their loan size, repayment speed, savings frequency, and income level in order to tailor financial products more effectively. Which data mining technique is most suitable for this task? Explain its purpose and benefits in this microfinance context. (3 marks)

Provide one other different realistic example of how NsambyaMic could use each of the following techniques: (i) Regression, (ii) Classification, (iii) Clustering. In each case, briefly describe the data used and the business goal. (4 marks each)

Suggest one appropriate machine learning algorithm for each of the following tasks and briefly explain why it is suitable: (i) Predicting the amount a client will save in the next three months, (ii) Classifying whether a loan application is low-risk or high-risk, (iii) Segmenting clients into behavioral groups based on their financial activity. (3 marks each)

Question 3:

[Figure from the original paper could not be retrieved]

style="width: 442px;" data-filename="figure1.JPG">

(i) What is the target attribute. Why ? (4 marks

(ii) Based on the nature of the target variable, determine whether this is a classification or regression problem. Justify your answer. (4 marks)

(iii) Identify the input features in the dataset and state how many features are used. (4 marks)

(iv) How many records (instances) are present in this dataset? (4 marks)

(v) List the possible outcomes (labels) that the model will predict. (4 marks

Question 4:

(a) With the aid of examples, define the following key concepts in data mining, clearly

highlighting their purposes and how they differ from each other: (5 marks each)

(i) Regression

(ii) Classification

(iii) Clustering

(b) For each of the three data mining tasks listed in part (a), name one common algorithms that are typically used. (1 mark each)

(c) Briefly explain the significance of data mining in modern data analysis and decision-making processes. (4 marks)

Question 5:

[Figure from the original paper could not be retrieved]

style="width: 305px;" data-filename="image.png">

(a) Identify whether the task of predicting car prices from the given dataset is a regression or classification problem. Justify your answer with reference to the nature of the target variable.

(b) Suggest a suitable data mining algorithm that can be applied to this dataset for predicting car prices. Explain why it is appropriate.

(c) Discuss two data preprocessing steps that should be carried out on this dataset before applying a machine learning model.

(d) If you were to perform a clustering task using only the numerical attributes, explain how you would group similar cars and what insights could be derived from such clusters.

Question 6:

(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 7:

(a) Briefly explain the following data warehouse applications (4 marks each)

(i) Information processing

(ii) Analytical processing

(iii) Data mining

(b) Explain the distinct features between OLTP (Online Transaction Processing) and OLAP (Online Analytical Processing) (8 marks).

Question 8:

For each of the following business-related problems, indicate whether the appropriate data mining technique is Classification, Regression, or Clustering. (Write either classification or regression, or clustering against each problem) (2marks each)

(i) Predicting future sales revenue based on seasonal trends and historical sales data.
(ii) Segmenting customers based on their buying behavior and demographics for targeted marketing.
(iii) Determining whether a credit card transaction is fraudulent.
(iv) Forecasting employee turnover rates based on performance and engagement data.
(v) Categorizing customer support tickets into topics such as billing, technical issue, or general inquiry.
(vi) Estimating the market value of a company based on its financial indicators.
(vii) Grouping similar products based on customer reviews and product features.
(viii) Predicting whether a customer will respond to a promotional campaign.
(ix) Estimating the delivery time for orders based on location, order size, and traffic data.
(x) Classifying business leads as high, medium, or low potential based on interaction history.

Question 9:

(a) What does K mean in the K-means algorithm clustering algorithm (3 marks)

(b) Draw a flowchart to illustrate steps in the K-means algorithm (10 marks)

(c) Explain each step in the K-means clustering algorithm (7 marks)

Question 10:

(a) Explain the difference between data, information, and knowledge in the context of data mining and business intelligence. Give one example of each. (12 marks)

(b) Explain any four data mining elements (8 marks)

Question 11:

(a) Detail five typical OLAP operations, explaining what each operation accomplishes in terms of data analysis. (10 marks)

(b) Explain the four key characteristics of a data warehouses (10 marks)

Question 12:

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

each)

(a) Data extraction:

(b) Data cleaning:

(c) Data transformation:

(d) Load:

(e) Refresh