

Kasarani Campus Off Thika Road Tel. 2042692 / 3 P. O. Box 49274, 00100 NAIROBI Westlands Campus Pamstech House Woodvale Grove Tel. 4442212 Fax: 4444175

# **KIRIRI WOMEN'S UNIVERSITY OF SCIENCE AND TECHNOLOGY UNIVERSITY EXAMINATION, 2024/2025 ACADEMIC YEAR** FOURTH YEAR, FIRST SEMESTER EXAMINATION FOR THE BACHELOR OF BUSINESS AND INFORMATION TECHNOLOGY KBI 2404 – DATA WAREHOUSES AND KNOWLEDGE MINING TECHNOLOGIES

Date: 05<sup>TH</sup> December 2024 Time: 8:30AM - 10:30AM

#### **INSTRUCTIONS TO CANDIDATES** ANSWER QUESTION ONE (COMPULSORY) AND ANY OTHER TWO QUESTIONS **QUESTION ONE (30 MARKS)**

- Giving examples discuss the following classification of knowledge mined in data mining system a)
  - Classification i)
  - ii) Prediction
  - iii) Clustering
  - iv) **Outlier** Analysis
- (4 Marks) b) Dataset represents a fictional set of customer information for a small retail store, including features such as customer ID, age, gender, annual income, and spending score, explain the steps you would take to perform knowledge discovery on this dataset. (5 Marks)
- Describe how decision trees handle missing values in the dataset and what strategies can be employed c) to address missing data during the training phase? (6 Marks)
- The assumptions of linear regression are critical for ensuring the validity and accuracy of the model's d) predictions. Describe two linear regression assumption (4 Marks)
- With example describe how blockchain technology can be applied to electronic money systems. e)
  - (5 Marks) (6 Marks)

Using R, solve the system of linear equations: f)

#### 2x+3y+z=1x+2y+3z=23x+y+2z=3

Show the R code and the solution.

## **QUESTIONS TWO (20 MARKS)**

- a) The Kenya Anti-Terrorism commission collects data about the potential attacks that may happen in the country. The commission is looking for an employee who can be able to assist in the mining of Knowledge from the large amount of data they have collected. As a candidate for the position, explain the knowledge discovery steps that they can follow to acquire useful information. (8 Marks)
- b) Decision Tree is a supervised learning method used in data mining for classification and regression methods.
  - i) Using a suitable example discuss why they are suitable in data mining. (6 Marks)
  - The decision tree algorithm is based on three parameters: D, attribute list, and Attribute ii) \_selection method. Explain the parameters. (6 Marks)

### **QUESTION THREE (20 MARKS)**

- Describe visualization and explain two data visualization techniques. a) (5 Marks) Consider this classification model – a group of students have been classified based on two attributes: b) educationLevel and scholarshipAmount. EducationLevel is further divided into two sub-attributes: undergraduate and postgraduate. ScholarshipAmount is also further sub-divided into two subattributes: 200,000 and 800,000. The reasoning that if educationLevel is undergraduate then the class is "BSc", and if educationLevel is postgraduate and scholarshipAmount is 200,000 then the class is "MSc", and if educationLevel is postgraduate and scholarshipAmount is 800,000 then the class is "PhD". You are required to represent this classification model in the following forms: If-then rules i) (5 Marks) ii) Decision tree (5 Marks) Neural network. (5 Marks) iii) **QUESTION FOUR (20 MARKS)** Give R expressions that return the following matrix and vectors. a) int {10:20} 10,11,12,13,14,15,16,17,18,19 i) D (3 Marks) values <- c(10,20,30,40,50) "The standard deviation is:" ii) (3 Marks) 15.81139 iii) "The mean is:" (3 Marks) 30
- b) Given a quadratic function =  $5x^2 + 4x 2$ , Write commands in R to compute the root of the expression. (5 Marks)
- c) Data Preprocessing is key in data mining, it ensures the quality and relevance of your data before beginning any data mining tasks explain three R packages used in data preprocessing (6 Marks)

## **QUESTION FIVE (20 MARKS)**

- a) As a software engineer at Kiriri Women's University you are tasked to design a data mining system to examine the university course database, which contains the following information: the name, address, and status (e.g., undergraduate or graduate) of each student, the courses taken, and their cumulative grade point average (GPA). Describe the architecture you would choose. What is the purpose of each component of this architecture? (6 Marks)
- b) Data warehouse architecture is the proper arrangement of the software and hardware components. With the aid of a diagram explain the components of data warehouse that suit the requirements of any organization for maximum benefit. (10 Marks)
- c) Give an example of a real-world application of an artificial neural network. Explain how the this network structured. (4 Marks)