



# MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya

Tel: +254(0) 799 529 958, +254(0) 799 529 959, + 254 (0) 712 524 293,

Website: [info@must.ac.ke](mailto:info@must.ac.ke) Email: [info@must.ac.ke](mailto:info@must.ac.ke)

---

## University Examinations 2024/2025

FIRST YEAR FIRST SEMESTER FOR THE DEGREE OF BACHELOR OF SCIENCE IN DATA  
SCIENCE

### CDS 3103: PRINCIPLES OF DATA SCIENCE

**DATE: DECEMBER 2024**

**TIME: 2 HOURS**

---

**INSTRUCTIONS:** Answer question *ONE* (Compulsory) and any other *TWO* questions

---

#### QUESTION ONE (30 MARKS)

- a) Define each of the following terms as used in Data Science;
- i. Data Science (1 Mark)
  - ii. Classification (1 mark)
  - iii. Clustering (1 Mark)
  - iv. Data Set (1 Mark)
- b) Traditional database (DB) consume a lot of time moving data from one DB to another. These challenges are overcome by the In-DB ML.
- i. Describe the In-DB ML concept as used in Data Science (2 Marks)
  - ii. Outline any four merits of using In-DB ML (4 Marks)
- c) Distance metrics play a significant role for both clustering and classification data science problems. Using appropriate examples, describe three different distance metrics that can be adopted in these problems (6 Marks)
- d) Compare and contrast the data science pyramid with the DIKW Eliot's pyramid (6 Marks)

- e) Using data science, different types of novel patterns can be extracted to inform decision making in an organization. Using appropriate examples, describe how each of the extracted patterns can guide each of the following tasks;
- i. Medical imaging (2 Marks)
  - ii. Time Series Analysis (2 Marks)
  - iii. Social Networking Analysis (2 Marks)
  - iv. Transportation System (2 Marks)

## QUESTION TWO (20 MARKS)

- a) Data Science is fundamentally dependent on data from which datasets are drawn. In its basic structure, data can be described as a set of values of qualitative or quantitative nature to abstract a real-world object.
- i. Define a real world object known as Student with ten (10) distinct attributes (4 Marks)
  - ii. For each of the above ten (10) attributes, indicate their respective data type using a table format as illustrated below; (5 Marks)

Attribute Name	<i>Data Type</i>

- iii. Create a 10 x 10 dataset of the above (i) defined object (5 Marks)
- iv. Categorizing the attributes into their respective data types, clearly indicate the probable data analysis techniques that can be applied to it. (6 Marks)

## QUESTION THREE (20 MARKS)

- a) Brain Tech company has been contracted by a leading financial institution to undertake a data science project to its enormous data dating back to 1950 when it was inceptioned. The data seeks to understand its business dynamics revolving around marketing, customer retention levels and profit making. As the lead Scientist, you are tasked with leading the Brain Tech team in undertaking this task. Required;
- i. The CRISP-DM model has been proposed as the guiding tool to undertake this exercise. Using an illustration, describe it (8 Marks)

- ii. In each of the stages explained in (i) above, detail the activities you would undertake in accomplishing your task as the lead data scientist in Brain Tech (6 Marks)
- iii. Besides CRISP-DM, describe other two techniques that would be adopted to address the problem at hand clearly stating their pros and cons (6 Marks)

#### **QUESTION FOUR (20 MARKS)**

- a) Using an illustration discuss the Data-Architecture for Data Science (10 Marks)
- b) Using an illustration, explain the key components of a simple neural network (4 Marks)
- c) Distinguish between each of the following terms as used in Neural Networks
  - i. Feed forward network and recurrent network (2 Marks)
  - ii. Threshold and bias (2 Marks)
  - iii. Pattern Classification and Pattern Association (2 Marks)

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

- a) Explain the importance of data pre-processing in machine learning (4 Marks)
- b) Differentiate between normalization and standardization (4 Marks)
- c) Describe two common data imputation techniques using in machine learning (4 marks)
- d) Perform feature scaling (normalization and standardization) on the following dataset and explain your chosen approach: (8 Marks)

Age: [25, 30, 35, 40, 45, 50, 55, 60]

Salary (in Kes1000): [35, 45, 55, 65, 75, 85, 95, 105]