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 Email: info@must.ac.ke

University Examinations 2024/2025

THIRD YEAR FIRST SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF COMPUTER TECHNOLOGY

FOURTH YEAR FIRST SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF DATA SCIENCE

CDS 3301: MACHINE LEARNING

DATE: JANUARY 2025 **TIME: 2 HOURS**

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

QUESTION ONE (30 MARKS)

a) Machine learning life cycle involves several major steps. List them. (5 marks)

b) Analyze any two Key differences between:

i. Artificial Intelligence (AI) and Machine learning (ML). (2 marks)

ii. Supervised and Unsupervised learning. (2 marks)

iii. Traditional programming and machine learning. (2 marks)

iv. Underfitting and Overfitting (2 marks)

c) Write a short explanation on the following unsupervised learning algorithms: (8 marks)

i. K-means clustering

ii. KNN (k-nearest neighbors)

iii. Neural Networks

iv. Learning rate

d) Describe the categories of the Ensemble Learning technique providing an example of each. (4 marks)

e) Discuss the steps in the design of a learning system. (3 marks)

f) Explain any two disadvantages of ML (2 marks)

QUESTION TWO (20 MARKS)

a) We have trained a machine learning model to identify whether the object in an image is a cat. Now we use 200 pictures to verify the model performance. Among the 200 images, objects in 170 images are cats, while others are not. The identification result of the model is that objects in 160 images are cats, while others are not. Calculate the following:

   i. Precision: P (2 marks)

   ii. Recall: R (2 marks)

   iii. Accuracy: ACC (2 marks)

b) Discuss the concepts of **"Responsible and Ethical Machine Learning"** (4 marks)

c) Describe the following: (4 marks)

   i. K-means clustering

   ii. Regularization

   iii. Naive Bayes

   iv. SVM

   v. Large language model

d) The following figure shows a classification when a decision tree is used. The classification result is impacted by three

attributes: Refund, Marital Status, and Taxable Income. Draw the Decision tree. (4 marks)

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125,000	No
2	No	Married	100,000	No
3	No	Single	70,000	No
4	Yes	Married	120,000	No
5	No	Divorced	95,000	Yes
6	No	Married	60,000	No
7	Yes	Divorced	220,000	No
8	No	Single	85,000	Yes
9	No	Married	75,000	No
10	No	Single	90,000	Yes

QUESTION THREE (20 MARKS)

a) Enumerate and explain the steps involved in building the Machine learning model. (4 marks)

b) Assume you have trained the model and obtained predictions. How would you evaluate the model's performance? Discuss at least two evaluation metrics and their significance in the context of this problem. (4 marks)

c) Discuss the factors that decide the selection of the machine learning algorithm. (3 marks)

d) Kenya Wildlife Services would like to implement biodiversity surveillance systems in the national parks. Evaluate and advice the suitability of the following feature selection methods: (9 marks)

i. Filter Method,
ii. Wrapper Method
iii. Embedded Method.

QUESTION FOUR (20 MARKS)

a) Describe the steps required during the workings of K-NN algorithm (4 marks)

b) Discuss using a suitable example, the difference between the following machine learning approaches:

i. Regression and Classification (4 marks)
ii. Supervised and Unsupervised Learning 4 marks)
iii. Precision and Recall (4 marks)

c) Analyze the pros and cons of why a **Cleaning Robot** company should consider using Reinforcement learning approach. (4 marks)

QUESTION FIVE (20 MARKS)

a) A local Research Lab is considering employing machine learning algorithms to predict crop pests based on various features. Briefly describe each of the following algorithms and discuss their potential application.

i. Random Forest Algorithm (5 marks)

ii. Decision Tree Algorithm (5 marks)

b) Define overfitting in machine learning. Why is it a problem, and how can it be addressed? Describe some techniques for preventing overfitting in machine learning models. (8 marks)

c) Analyze the difference between Parameters and Hyperparameters in Models. (2 marks)