



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

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

CCD 7104: MACHINE LEARNING

DATE: DECEMBER 2024

TIME: 3 HOURS

INSTRUCTIONS: *Answer any three questions*

QUESTION ONE (20 MARKS)

- a) "Much of what we do with machine learning happens beneath the surface. Though less visible, much of the impact of machine learning will be of this type — quietly but meaningfully improving core operations. " **JEFF BEZOS**. Using examples of applications in the Kenyan context, briefly explain the validity or otherwise of the statement. (4 Marks)
- b) Using examples, briefly explain the following in relation to design of intelligent systems:
- Expert System
 - Case-based learning
 - Fuzzy Logic
 - Gradual Rules
 - Web mining
- (5 Marks)
- c) Using examples, differentiate between the following approaches in relation to design of intelligent agents.
- Clustering and Semi-supervised learning
 - Supervised learning and Reinforcement learning
 - Pattern Classification and Pattern Association
- (6 Marks)
- d) "A few weeks ago, Stephen Hawking opened the world's eyes to the dangers of Artificial Intelligence (AI), warning that it has the potential of outsmarting humans in the financial markets." [futuristspeaker, com, 2016]. Using real-life examples, discuss the validity or otherwise of the

above statement and hence differentiate between the natural human decision process (biological) and the intelligent agent (artificial intelligence) decision making process. (5 Marks)

QUESTION TWO (20 MARKS)

- a) Although KNN has been shown to perform fairly well as a classifier, it has several limitations. Using examples, briefly discuss how KNN works, its limitations and how they can be overcome. (5 Marks)
- b) A research was carried out aimed at classifying customers in Taiwan as either- credible or not credible clients. A sample of the Attributes X1, X12, X13 and outcome (whether the customer defaulted payment next month or not) as shown on **Table 1**. X1: Amount of the given credit (NT dollar), X12 = amount of bill statement in September, 2005; X13= amount of bill statement in August, 2005.

Table 1

ID	X1	X12	X13	Y
	LIMIT_BAL	BILL_AMT1	BILL_AMT2	default payment next month
1	4	4	10	Yes
2	4	8	5	No
3	9	6	10	Yes
4	8	8	6	No
5	5	8	3	Yes
6	1	4	7	No
7	10	8	6	Yes
8	2	4	2	No
9	2	6	3	Yes
10	7	3	3	No
11	4	1	9	Yes
12	5	2	7	No

- i) Suppose that a new customer walks in with the following characteristics (X1=5, X12=4, X13=6). Using K nearest Neighbor (K=3), make the prediction for the type of customer using the data set. (7 Marks)

- ii) Using the data set above and 6 Cross validation, determine the accuracy of KNN (K=3)
(8 Marks)

QUESTION THREE (20 MARKS)

- a) Suppose you have been asked to develop a "1- Life: a life Insurance expert system to value the premiums to be paid by a customer". Using diagrams if necessary, briefly describe the basic components and working of such an intelligent system. (5 Marks)
- b) Car Evaluation Database is provided below on Table 2 to derive a simple hierarchical decision model¹. The model should evaluate cars according: unacceptable (unacc), acceptable (acc), good and very good

(v good).

Table 2

Car	buying	maint	safety	Class
1	med	med	high	vgood
2	med	med	high	vgood
3	med	low	high	vgood
4	med	low	high	vgood
5	vhigh	vhigh	med	unacc
6	vhigh	vhigh	high	unacc
7	vhigh	vhigh	low	unacc
8	med	low	med	good
9	med	low	med	good
10	low	vhigh	low	unacc
11	low	vhigh	med	unacc
12	low	low	low	good

Required:

- i. Using Gini Index as impurity measure, construct a decision tree for the intelligent agent to do car evaluation. (10 Marks)
- ii. Write simple code segment in C++ or Java programming language to implement an algorithm of the above decision tree (5 Marks)

QUESTION FOUR (20 MARKS)

- a) Suppose that you have the data below on Table 3 and you wish to group the customers into two clusters for better client focus. Use K-Means to derive the clusters. (7 Marks)

Table 3

ID	X1	X12	X13
	Items Purchased	Minimum total Amount	Maximum total Amount
1	5	2	5
2	5	2	7
3	5	7	7
4	9	5	9
5	7	8	9

- b) Suppose you have been given the following Data of purchases made by 10 customers in a Wakulima supermarket as shown on Table 4.

Table 4

Transaction	Items			
Transaction 1	Chicken	Colgate	Milk	
Transaction 2	Milk	Colgate		
Transaction 3	Omo	Bread	Elianto	Colgate
Transaction 4	Bread	Milk		
Transaction 5	Maize Flour	Geisha	Milk	Sukumawiki
Transaction 6	Milk	Maize Flour	Elianto	Colgate Chicken
Transaction 7	Chicken	Milk	Colgate	Omo
Transaction 8	Colgate	Sukumawiki	Maize Flour	Bread
Transaction 9	Chicken	Elianto	Sukumawiki	
Transaction 10	Elianto	Maize Flour	Chicken	

i) Using a support of 25%, extract the sets of items that are frequently purchased. (8 Marks)

ii) Write simple Java/C++ algorithm code segment to implement the above association rules (5 Marks)

QUESTION FIVE (20 MARKS)

a) Using a single layer Artificial Neural Network for an **AND gate** as an example

i) Describe the architecture and the computational task of the artificial neural network. (5 Marks)

ii) Briefly discuss the limitations of artificial neural networks and how they can handled. (3 Marks)

b) Suppose that you have been asked by a bank to design an intelligent system to distinguish between three populations distributions (Z, P and N) so as to determine the diseases that are likely to be most prevalent in the area. You have been provided by the patterns below on Figure 1.

Z				P				N			
#	#	#	#	#	#	#	#	#	.	.	#
.	.	#	.	#	.	.	#	#	#	.	#
.	#	.	.	#	#	#	.	#	.	#	#
#	#	#	#	#	.	.	.	#	.	.	#

i) Use the three patterns above (Z, P and N) to compute the weights to be used to recognize of the three population distributions. (7 Marks)

ii) Use the constructed artificial neural network in i) above to classify the distribution shown on Figure 2 as either Z, P or N. (5 Marks)

Figure 2

#	.	.	#
.	#	#	.
.	#	#	.
#	.	.	#