# **Question Paper**

Exam Date & Time: 02-Jan-2023 (10:00 AM - 01:00 PM)



## MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal

First Semester Master of Engineering - ME (Artificial Intelligence & Machine Learning) Degree Examination - January 2023

#### Applied Machine Learning [AML 5102]

#### Marks: 100

Duration: 180 mins.

#### Monday, January 1, 2023

### Answer all the questions.

- 1)
   1. Describe how Machine Learning can help the institution by giving specific examples of how (10) techniques, such as clustering, classification, regression, and anomaly detection can be applied.
   (10)

   [L2] [TLO 1.1] [1 x 10 =10 Marks]
   (10)
- 2) Given the following dataset where color, age and weight are the input features and category the (10) target feature label. The details of the features are given below.COLORpossible values:Red(R), Green (G), Blue (B)
   AGEpossible values:Young(Y), Old (O)

WEIGHTpossible values:Light (L), Medium (M), Heavy (H)

Color	Age	Weight	Category
R	0	Н	+
В	Y	L	+
G	Y	L	+
R	Y	Н	+
G	0	L	-
G	Y	L	-
В	0	Н	-

Apply Hamming distance. What predication the KNN will return for color =B, Age =O and Weight=L with K=3 using the dataset given above.

What is the difference between parametric and non-parametric models. [L4,L3] [TLO1.2] [5+5 =10 Marks]

The classifier made 500 predictions for the presence of cancer. Among 75 were confirmed positive (10) cases, 125 false-positive cases, 50 false negatives cases, and 250 true negatives.

- a. Develop a the confusion matrix
- b. Identify True Positive, False Positive, True Negative, and True Positive.

- c. Calculate Accuracy and Misclassification Error.
- d. Calculate Specificity and False Positive Rate.

e. Calculate Precision and Recall.[L6] [TLO 2.1] [2 x 5=10 Marks]

a. What are training, validation test data sets, and how are they used differently? What proportion of (10) data will you allocate for the training, validation, and test sets?

b. Consider the following dataset with one real-valued input x and one binary output 'y'. we are going to use k-NN with unweighted Euclidean distance to predict 'y' for 'x'. What is the leave one out cross validation error of 1-NN on this dataset? Give your answer as the number of misclassifications?

х	У
0.1	-
0.6	+
1	+
1.6	-
2	+
2.5	+
3.2	-
3.5	-
4.1	+
4.9	+

[L3] [TLO 1.2] [5 x 2 = 10 Marks]

5a. Draw Decision Tree for the given below datasets using ID3.

(10)

Gender	Major	Like
Male	Maths	Yes
Female	History	No
Male	Computer Science	Yes
Female	Maths	No
Female	InformationScience	Yes
Male	InformationScience	Yes
Male	History	No
Female	Maths	No

b.Explain any two Challenges faced by Decision Tree and provide solutions to overcome.[L4, L3]

5)

	Α	В	С	D
Α	0	1	4	5
В		0	2	6
С			0	3
D				0

[L4][TLO2.2] [1X10=10 Marks ]

7. Compare bagging and Boosting with respect to

(10)

(10)

- Similarities

- Differences[L3] [TLO3.2] [1X10=10 Marks]

8. Why do we need Principle Component Analysis(PCA)algorithm for analysing the data ? Explain (10) the role of Eigen vector and Eigen value in PCA algorithm.
L3 (TLO 2.2) [1 X10 = 10 Marks]

9. For a set of data points, 3 different classifiers are built with different k values. Which classifier (10) you will select and why?

When k =2



When k=5





7)

8)

9)



a. What happens if you select the lowest k value?

b. What happens if you select the highest k value?[L4,L3][TLO2.1] [ 5 + 5 = 10 Marks ]

10. Consider the following dataset for a binary classification problem:

(10)

X <sub>1</sub>	X <sub>2</sub>	Y
1.0	-1.0	1
2.0	3.0	-1
2.5	1.5	-1
1.5	2.0	1
2.0	-4.0	1

Calculate the margins of these samples from the hyperplane -x1+x2+1 = 0. Which sample has the largest positive margin, and which one the largest negative margin?

-----End-----

10)