A

Reg. No. :

Answer ALL Questions

PART A - (10 x 1 = 10 Marks)

1. The categories in which Machine learning approaches can be traditionally categorized are _____   CO1- U

   (a)Supervised Learning   (b) Unsupervised   Learning

   (c)Reinforcement Learning   (d) All of the Above

2. _____ is the machine learning algorithms that can be used with labeled data   CO1- U

   (a) Regression Algorithms   (b) Clustering Algorithms

   (c) Association Algorithms   (d) All of the Above

3. _____ is the scenario when the model fails to decipher the underlying trend in the input data.   CO1-U

   (a) Over fitting   (b) Under fitting   (c) Both (a) and (b)   (d) None of the Above

4. The supervised learning problems can be grouped as_____.   CO1-U

   (a) Regression Problems   (b) Classification Problems

   (c) Both (a) and (b)   (d) None of the Above

5. In K-Means clustering, the "K" represents: CO1-U

   (a) The number of features   (b) The number of clusters

   (c) The number of iterations   (d) The number of data points

6. Which clustering algorithm is most suitable for datasets with clusters of arbitrary shape and noise?   CO1-U

    (a) K-Means   (b) DBSCAN

    (c) PCA   (d) Agglomerative Clustering

7. Which of the following is an example of a categorical variable?   CO1- U

    (a) Age   (b) Height

    (c) Blood group   (d) Temperature

8. The process of converting continuous variables into discrete intervals is called   CO1- U

    (a) Scaling   (b) Normalization   (c) Binning   (d) Encoding

9. Fault detection in electrical systems is mainly used to: CO1- U

    (a) Increase voltage   (b) Identify abnormal operating conditions

    (c) Reduce battery size   (d) Speed up charging

10. Predictive maintenance helps in   CO1- U

    (a) Fixing equipment only after failure   (b) Predicting and preventing failures before occur

    (c) Reducing the number of sensors   (d) Increasing downtime

PART – B (5 x 2= 10 Marks)

11. Define Machine Learning and elucidate its fundamental principles.   CO1- U

12. You have the following points: (1, A), (3, B), (5, B), (6, A). Using k = 3, classify the new point by KNN at x = 4.   CO2 - App

13. Given the 1D data points: {2, 4, 10, 12}, and k=2, initialize cluster centroids as 2 and 10. Perform one iteration of K-Means clustering   CO3 - App

14. Differentiate between binning and discretization, providing one practical application for each.   CO1 - U

15. Summarize the role of predictive maintenance in minimizing equipment downtime.   CO5 - U

PART – C (5 x 16= 80Marks)

16. (a) Illustrate the installation of Python and configure an ML environment using pip and necessary packages.Also Explain how to load and explore the Iris dataset using scikit-learn. Write Python code to display feature names, data shape, and target labels.   CO1 - U   (16)

Or

R3511

(b) Explain the types of Machine Learning with suitable examples. CO1 - U (16)
Compare supervised, unsupervised, and reinforcement learning in terms of data usage and output.

17. (a) Build a simple binary decision tree to predict Pass (Yes/No) based CO2 - App (16)
on Study Hours, using entropy and information gain.

SAMPLE	STUDY HOURS	PASS
1	1	NO
2	2	NO
3	3	YES
4	4	YES
5	5	YES
6	6	YES
7	7	NO
8	8	NO
9	9	YES

Or

(b) Develop a methodology to predict the value of Y in terms of X1 CO2 - App (16)
and X2 based on the given data use KNN approach with
K=3.Using the methodology predict the value of Y for X1=15.2
and X2=33.1

Sample	A	B	Y
1	11.35	23	Blue
2	11.59	22.3	Blue
3	12.19	24.5	Blue
4	13.23	26.4	Blue
5	13.51	30.2	Blue
6	13.68	32	Red
7	14.78	31.2	Red
8	15.11	33	Red
9	15.55	25.2	Red
10	18.85	39.9	Red

18. (a) Given the following data, use Principal Component Analysis to CO3 – App (16)
reduce the dimension from 2 to 1.

Feature	Example1	Example2	Example3	Example4
x	4	8	13	7
y	11	4	5	14

Or

R3511

(b)   Cluster the following data points using K means into three   CO3 – App   (16)
clusters, where the point are A1(2,10), A2(2,5),A3(8,4), B1(5,8),
B2(7,5), B3(6,4), C1(1,2), C2(4,9).

19. (a)   A dataset has a categorical feature Color with values   CO3 - Ana   (16)
$\{Red, Blue, Green, Red, Green\}$ and a target variable (price):
[100,150,200,120,180]
- Encode the variable using Label Encoding and One-Hot Encoding.
- Analyze the correlation of encoded values with Sales.

Or

(b)   Given ages of students = {5, 12, 15, 18, 24, 30, 40, 55, 60}:   CO3 - Ana   (16)
- Apply Equal-Width Binning into 3 bins and Equal-Frequency Binning into 3 bins.
- Evaluate the variance within bins for each method.

20. (a)   Analyze a voltage signal dataset using a Decision Tree Classifier   CO5 - Ana   (16)
based on fundamental and 3rd harmonic values. Determine the
best root split via entropy and information gain. Use the model to
classify the point (0.8,0.08) and explain the reasoning

Sample	Fundamental (pu)	3rd Harmonic (pu)	Class
1	1.0	0.05	Normal
2	0.7	0.10	Sag
3	1.2	0.15	Swell

Or

(b)   Daily load data (MW) for a substation is given: CO5 - Ana   (16)
[120,135,150,145,160,175,190] construct and interpret a linear
regression model to analyze the relationship between time (days)
and power demand. Analyze the model to estimate the load for
day 8. Critically evaluate the limitations of using linear regression
for forecasting in the context of power system variability.

R3511