

Duration 3 Hours

[Maximum Marks 80

NOTE: -1) Question 1 is compulsory.

2) Solve any three from the remaining five questions.

3) Assume suitable data if necessary.

4) Figures to the right indicate full marks.

- Q.1.** Attempt any four **20**
- Explain the concept of entropy in the context of decision trees.
  - Why is it important to preprocess the data before the development of a machine learning model?
  - What distinguishes unsupervised learning from supervised learning with respect to the presence of labeled data?
  - What is variance in the context of machine learning?
  - What is a loss function in machine learning? Why is a loss function necessary in machine learning?

- Q.2.** a) Explain Dimensionality reduction techniques. **10**  
 b) Discuss in detail steps involved in K means clustering with one example. **10**

- Q.3.** a) Explain how machine learning is used in healthcare applications. **10**  
 b) What is the role of support vectors in SVM? Explain the concept of a hyper plane in the context of SVMs. How does an SVM handle non-linearly separable data? **10**

- Q.4.** a) Explain the confusion matrix and its components. **10**  
 b) Given below is the confusion matrix for binary classification. Find Precision, Recall, Accuracy, False Positive Rate, True Negative Rate **10**

	Predicted No	Predicted Yes
Actual No	50	10
Actual Yes	5	100

- Q.5.** a) How are decisions represented in a decision tree? What is the purpose of the root node in a decision tree? Explain pruning. **10**  
 b) Explain Bayesian Belief Network. **10**

- Q.6.** a) What is linear regression, and what type of problem does it address? Explain different error measures used for performance of regression **10**  
 b) How is the regression line determined in linear regression? Find the linear regression Model of y in terms of x for the following dataset. **10**

x	1	2	3	4	5
y	1	2	1.3	3.75	2.25

\*\*\*\*\*