

Duration: 3hrs

Max Marks:80

- N.B. :** (1) Question No 1 is Compulsory.
 (2) Attempt any three questions out of the remaining five.
 (3) All questions carry equal marks.
 (4) Assume suitable data, if required and state it clearly.

Q1 Attempt any **FOUR** from the following [20]

- A Explain any five business applications of Machine learning.
- B What is dimensionality reduction? Explain how it can be utilized for classification and clustering task in Machine learning.
- C Explain performance evaluation metrics for binary classification with suitable example.
- D Explain Gini index along with an example.
- E Explain the concept of k fold cross validation.

Q2 A Write a short note on issues in Machine Learning. [10]

- B Compare Bagging and Boosting with reference to ensemble learning. Explain how these methods help to improve the performance of the machine learning model. [10]

Q3 A Consider the example below where the mass, y (grams), of a chemical is related to the time, x (seconds), for which the chemical reaction has been taking place according to the table. Find the equation of the regression line. Also explain performance evaluation measures for regression. [10]

Time, x (seconds)	5	7	12	16	20
Mass, y (grams)	40	120	180	210	240

- B What is Density based clustering? Explain the steps used for clustering task using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. [10]

Q4 A Explain Clustering with minimal spanning tree along with example. [10]

- B Consider the dataset given below with 3 features Color, Wig, Num. Ears and one output variable Emotion [10]

Color	G	G	G	B	B	R	R	R	R
Wig	Y	N	N	N	N	N	N	N	Y
Num. Ears	2	2	2	2	2	2	2	2	3
Emotion	S	S	S	S	H	H	H	H	H

- i) Find root node of decision tree using GINI index
- ii) Explain techniques can be used to handle over fitting in decision trees?

Q5 A Consider the use case of Email spam detection. Identify and explain the suitable machine learning technique for this task. [10]

- B Explain the Dimensionality reduction technique Linear Discriminant Analysis and its real-world applications. [10]

Q6 A Define following terminologies with reference to Support vector machine: Hyper plane, Support Vectors, Hard Margin, Soft Margin, Kernel [10]

- B Explain Ensemble learning algorithm Random Forest and its use cases in real world applications. [10]
