

RAMAKRISHNA MISSION VIDYAMANDIRA
(Residential Autonomous College affiliated to University of Calcutta)

B.A./B.Sc. SIXTH SEMESTER EXAMINATION, MAY 2025

THIRD YEAR [BATCH 2022-25]

COMPUTER SCIENCE [Honours]

Paper : DSE 3

Date : 10/05/2025

Time : 11 am – 1 pm

Full Marks : 50

Answer **any five** questions of the following :

[5×10]

1. a) You are given purchase patterns from a retail store. Apply k-means clustering ($k = 3$) on the data and interpret the result.

C-ID	Visit per month	Average card value (Rs.)
C1	1	1200
C2	7	700
C3	2	1000
C4	5	2000
C5	3	1500
C6	4	1400
C7	6	1000
C8	5	1200

- b) Explain how Federated Learning differs from traditional centralized machine learning. Why is Federated Learning considered privacy-preserving? [7+3]

2. a) Use logistic regression to predict whether a person will subscribe for LIC. Estimate the coefficients using gradient descent (one epoch is enough).

Person	Age	Income per month (₹)	Subscribed
A	30	50,000	Yes
B	38	60,000	Yes
C	25	30,000	No
D	40	60,000	Yes
E	28	35,000	No

- b) Define cost function and its role in logistic regression. [7+3]

3. a) You're provided with a sensor dataset from a greenhouse. Apply Decision Tree logic to classify if water is needed (Yes or No)

Record	Soil Moisture (%)	Temp (°C)	Humidity (%)	Water Needed
1	35	30	70	No
2	20	35	60	Yes
3	40	28	75	No
4	18	32	50	Yes

- b) Mention two real-world applications of decision trees beyond agriculture. [7+3]

4. a) The following data shows equipment conditions and whether a failure occurred. Build a simple Naive Bayes classifier and predict failure for a new record: (Vibration: High, Heat: High, Noise: Medium).

Record	Vibration	Heat	Noise	Failure
1	Low	Low	Low	No
2	High	High	High	Yes
3	Medium	Low	Medium	No
4	High	High	Medium	Yes
5	Low	Medium	Low	No

- b) What assumptions does Naive Bayes make, and when can it fail? [7+3]
5. a) Apply PCA step-by-step on the following patient dataset (2D projection). Find the direction of maximum variance.

Glucose	Blood Pressure
90	60
85	65
100	70
95	68
87	63

- b) Explain the difference between PCA and LDA. [7+3]
6. a) Classify the following tweet as Positive or Negative using KNN with $k=3$. New Tweet has Length = 18 words and #Emojis = 3.

Tweet	Length	Emojis	Sentiment
T1	20	2	Positive
T2	15	4	Negative
T3	25	3	Positive
T4	17	5	Negative
T5	19	3	Positive

- b) What happens when k is too low or too high? [7+3]
7. a) Define cross-validation and explain how k -fold cross-validation works. Compare it with leave-one-out cross validation (LOOCV) in terms of bias and variance.
- b) Explain how lack of explainability in black-box models can lead to ethical concerns and affect user trust in real-world AI systems. [7+3]

————— × —————