BE/Sem VII/Comp - CSE/C-scheme/NOV-25/10-11-25

Duration: 3hrs **[Max Marks:80]**

KONKAN GYANPEETH COLLEGE OF ENGINEERING KARJAT Dist. Raigad

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. Solve any **four** from following. [20]

a. Explain the distance metrics used in clustering.

b. Explain Regression line, Scatter plot, Error in prediction and Best fitting line.

c. What is the significance of optimal separating hyper plane in SVM?

d. Suppose that you were analysing data related to a group of 50 people applying for a grant. Each grant proposal was read by two readers and each reader either said "Yes" or "No" to the proposal. Calculate Precision, Recall, Accuracy

		Reader B (Predicted)	
		Yes	No
Reader A (Actual)	Yes	20	5
	No	10	15

e. Explain the concepts of entropy and Gini.

Q2. a. Explain any five applications of Machine Learning. [10]

b. Explain simple linear regression. Find regression line Y=aX+C using least square method and estimate the expenditure of company in 6th month using line as a model. The expenditure of an organization (in thousand) for every month is shown below, [10]

X(Month)	**1**	**2**	**3**	**4**	**5**
Y(Expenditure)	**12**	**19**	**29**	**37**	**45**

Q3. a. Explain kernel functions of SVM along with an example. [10]

b. Explain LDA along with an example. [10]

Q4. a. Explain the Random Forest algorithm in detail. [10]

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

Q5. a. Describe Multiclass classification. [10]

b. Explain DBSCAN clustering algorithm along with an example. [10]

Q6. Write detailed note on following. (**Any two**) [20]

a. XG Boost method of ensemble learning

b. Expectation Maximization Algorithm

c. Logistics Regression

****

muquestionpapers.com