U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

October 2024 Supplementary Examinations

Programme: B.E. **Semester: VI**
Branch: Information Science and Engineering **Duration: 3 hrs.**
Course Code: 22IS6PCMLG **Max Marks: 100**
Course: Machine Learning

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	CO	PO	Marks
1	a)	Illustrate the following: • Supervised and Unsupervised learning • Online and batch learning.	CO1		6+4
	b)	Apply the concept of data pre-processing and write the complete python code to transform the raw data into an understandable for the following dataset.	CO2	PO1	10

Student Name	Semester	Total_Marks	Grade
Anup	2	592	S
Ajit	2	560	A
Rohit	1	500	B
Ram	2	540	A
Ajay	3	NAN	S
Shruti	3	450	C
Joy	NAN	400	D
Anitha	1	542	A
Shrivatsa	2	500	B

		OR			
2	a)	Explain the following with appropriate python code • Grid-Search CV • Cross-validation & Correlations	CO1		5+5
	b)	i. Imagine a dataset with two independent features where one of the feature is numerical and the other one is categorical feature. Few of the values in the dataset are missing. Illustrate the different ways of fixing these missing values and write the appropriate python code for the same. ii. What is pipeline? Create a pipeline for the description given in 2b (i).	CO2	PO1	3+3
	c)	Explain any two challenges faced by machine learning algorithm.	CO1		04

			UNIT - II			
	3	a)	Illustrate multiclass classification with suitable example.	***CO1***		**08**
		b)	What is Gradient Descent? Apply the concept of batch gradient descent for X= 0.5, 2.3, 2.9 & Y=1.4,1.9,3.2 and find the values of intercept with slope as 0.64 for 2 iterations.	***CO2***	***PO1***	**08**
		c)	Justify your answer whether to use high precision or high recall for the following scenario: i) Suppose we have to build a model which detects if a video is safe for kids or not. ii) Say we have to build a model which detects shoplifters on the basis of surveillance image.	***CO2***	***PO1***	**04**
			OR			
	4	a)	A student dataset has four columns with three independent columns that includes marks of three subjects and a dependent variable column includes the overall average. i) For this dataset, how do we decide whether to apply the concept of linear regression or polynomial regression. Justify your answer. ii) Write the complete python code for the same with suitable performance measures. **Assume** – data is pre-processed	***CO3***	***PO2***	**3+7**
		b)	IRIS dataset has four independent features – petal length and width, sepal length and width that contains numerical values of around 150 instances and a dependent column as species – Iris setosa, Iris virginica and Iris versicolor. Write the complete python code that includes necessary pre-processing on independent features, logistics regression model for multiple classes, prediction on test data and performance measures.	***CO3***	***PO2***	**10**
			UNIT - III			
	5	a)	Design a decision tree classifier model using IRIS dataset. Write python code to plot the decision tree, performance measures and predict the class for the values [5,1.5].	***CO3***	***PO2***	**10**
		b)	What are decision trees? Generally, decision tree overfits the data, justify which hyper parameters stops the tree from overfitting with suitable examples.	***CO2***	***PO1***	**05**
		c)	Illustrate the following with suitable formulas: • Gini index • CART	***CO1***		**2+3**
			UNIT - IV			
	6	a)	What is curse of dimensionality? Illustrate two main approaches of dimensionality reduction.	***CO2***	***PO1***	**08**

		b)	i) When should I use ensemble learning techniques and what are the most popular ensemble methods. ii) Explain with any two ensemble methods along with python code.	***CO2***	***PO1***	**2 + 10**
			UNIT - V			
	7	a)	Apply K(=2) Means algorithm over the data (185, 72), (170, 56), (168, 60), (179,68), (182,72), (188,77) up to two iterations and show the clusters. Initially choose first two objects as initial centroids.	***CO2***	***P1***	**06**
		b)	Write the python code for clustering the data points given in 7(a) using k-means clustering. Use the cluster number that you have got from 7(a) and plot the elbow method and print silhouette score.	***CO3***	***PO2***	**08**
		c)	What is DBSCAN and explain the min-points, core points, epsilon terminologies.	***CO1***		**06**

****