U.S.N.										

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

Autonomous Institute Affiliated to VTU

February / March 2023 Semester End Main Examinations

Programme: B.E. **Semester: V**
Branch: Information Science and Engineering **Duration: 3 hrs.**
Course Code: 20IS5PCMLG **Max Marks: 100**
Course: Machine Learning **Date: 21.02.2023**

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

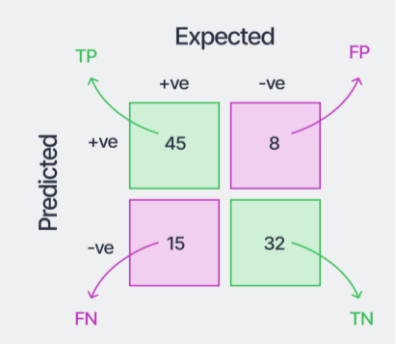
1	a)	Briefly explain any two types of Machine learning systems.	**06**
	b)	Describe in short the steps involved in an end-to-end machine learning project.	**10**
	c)	Why is the training set larger than the test set in machine learning? What is the purpose of the validation set?	**04**

OR

2	a)	Define Machine learning. Give any two real life examples where you can use machine learning.	**05**
	b)	i. Describe the various ways to discover and visualize the data to gain insights with sample code. ii. Why is data cleaning prominent in any machine learning project? How do you handle missing values in the dataset?	**10**
	c)	Explain Grid Search and Randomized Search hyperparameter tuning techniques.	**05**

UNIT - II

3	a)	Find the precision, recall and F1 score for the confusion matrix given below:(write the formulas)	**06**



b) In a given dataset, there are n features(numerical) and one target(categorical). The target has n classes. Write a python code using sklearn libraries to load the dataset(data.csv), preprocess the dataset, split the dataset and train a binary classifier to determine if the given new instance belongs to a particular class or not along with classification report. **10**

c) A data scientist has trained the linear regressor, he wants to reduce the cost function. Identify the technique he can apply and how does that technique work? **04**

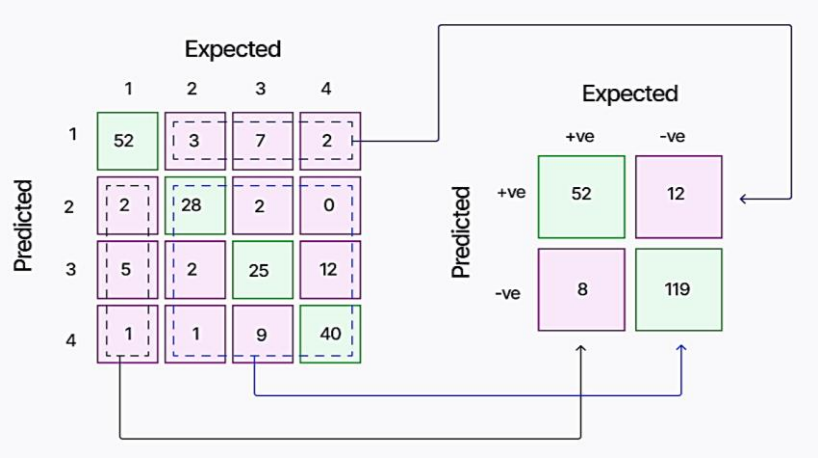
OR

4 a) Briefly explain the different performance measures of a regressor. **05**

b) Find the coefficients of the linear regression model using Ordinary least squares for the given data points. Predict Y when X={8, 12} **10**

X	Y
0	1
3	5
4	6
7	8
10	9

c) Identify and elicit the type of classification. **05**



UNIT - III

5 a) Illustrate the working of Decision tree algorithm for the following sample dataset. Draw the decision tree based on Gini scores for the following instances given as a training set. **10**

Past Trend	Open Interest	Trading Volume	Return
Positive	Low	High	Up
Negative	High	Low	Down
Positive	Low	High	Up
Positive	High	High	Up
Negative	Low	High	Down
Positive	Low	Low	Down
Negative	High	High	Down
Negative	Low	High	Down
Positive	Low	Low	Down
Positive	High	High	Up

b) Differentiate between Gini index and Entropy. List the hyperparameters of a Decision tree model. **3+2**

c) Identify and write any two scenarios where the decision tree model is a better model to train and justify. **05**

UNIT - IV

6 a) Briefly explain any two ensemble methods. **05**

b) Explain the working of Adaboosting by considering sample example. **10**

c) Justify the need of dimensionality reduction for MNIST dataset. How is PCA used in this context? **05**

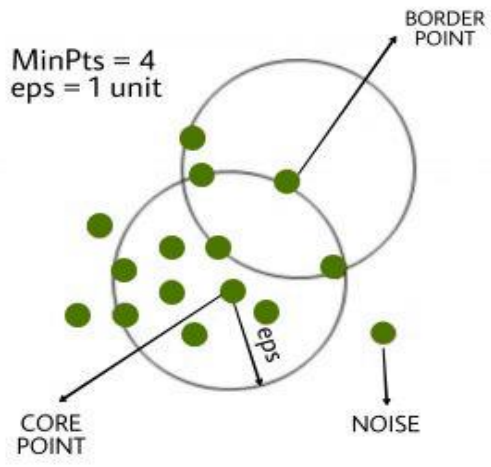
UNIT - V

7 a) When do you apply Unsupervised Learning? Give any one real life example where unsupervised learning can be used. **04**

b) Write the steps involved in finding k clusters using K-Means algorithm for the below dataset and write a python code to illustrate K-Means by finding the optimal k value and print the k centroids. **10**

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

c) Describe the labels in the figure given below w.r.t DBSCAN. When is DBSCAN used instead of the K-Means model? **06**



******