U.S.N.										

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

Autonomous Institute Affiliated to VTU

September / October 2023 Supplementary Examinations

Programme: B.E. **Semester: V**
Branch: Information Science and Engineering **Duration: 3 hrs.**
Course Code: 20IS5PCMLG **Max Marks: 100**
Course: MACHINE LEARNING **Date: 27.09.2023**

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

UNIT - I

1 a) With suitable diagrams distinguish how a spam filter model using machine learning technique is better compared to a traditional programming technique. **10**

b) Design the following steps involved in data pre-processing using Sklearn libraries considering Health Dataset. **10**

Ethnicity	Height (CM)	Weight (Kg)	Will survive till 70
White	186	90	Yes
African	185	98	No
White	175	80	No
White	180	88	Yes
Asian	178		No
Asian	172	72	Yes
African	178	75	No
White		89	Yes
African	186	90	Yes

- Import Libraries
- Importing the dataset
- Splitting the dataset into the Training set and Test set
- Taking care of missing values
- Encoding the Independent Variable
- Encoding the dependent Variable
- Feature Scaling -Apply Standard Scalar on complete X-train and X-test

OR

2 a) Design the following steps involved in data pre-processing using Sklearn libraries considering the Region Dataset. **10**

- Import the libraries
- Import the Dataset
- Handling of Missing Data
- Handling of Categorical Data
- Splitting the dataset into training and testing datasets

- Feature Scaling-Apply Standard Scalar on complete X-train and X-test

1	Region	Age	Income	Online Shopper
2	India	49	86400	No
3	Brazil	32	57600	Yes
4	USA	35	64800	No
5	Brazil	43	73200	No
6	USA	45		Yes
7	India	40	69600	Yes
8	Brazil		62400	No
9	India	53	94800	Yes
10	USA	55	99600	No
11	India	42	80400	Yes

b) Illustrate the following with suitable examples and diagrams: **10**
- Supervised
- Unsupervised
- Semi-Supervised
- Encoding techniques

UNIT - II

3 a) Design a binary classifier model considering MNIST dataset and prove the performance of the classifier by considering accuracy, confusion matrix, precision and recall as performance measures using Sklearn libraries. **10**

b) What is Gradient Descent? Explain the Batch Gradient Descent with suitable example. **10**

OR

4 a) Design the following steps involved in building linear regression model considering the House-price dataset. **10**
- Import Libraries
- Importing the dataset
- Splitting the dataset with test size=30%
- Feature Scaling -Apply Standard Scalar
- Training the Linear Regression model on the Training set
- Print performance measure MAE,RMSE

b) Design the steps involved in building the Logistics Regression model using Social_Network_Ads dataset. **10**

Note: Split the dataset with test size =25% and apply Standard Scalar and print the confusion matrix, accuracy score and classification report.

UNIT - III

5 a) Design the following steps for Decision tree classifier using Social_Network_Ads.csv **10**

Age	EstimatedSalary	Purchased
19	19000	0
35	20000	0
26	43000	0
27	57000	0
19	76000	0
27	58000	0
27	84000	0
32	150000	1

.

- Importing the libraries
- Importing the dataset
- Splitting the dataset with test size =25%
- Feature Scaling - Apply Standard Scalar
- Training the model
- Predicting a new result for 30,87000
- Print the Confusion Matrix, Classification report

b) Explain the CART algorithm along with cost function used for classification. **05**

c) With suitable examples, explain the different regularization hyper parameters used for restricting the shape of the decision tree. **05**

UNIT - IV

6 a) Explain the following with suitable examples: **10**

- Bagging
- Pasting
- Random Forests
- Extra Trees

b) Design the steps involved in PCA to reduce the dimensionality of MNIST dataset with variance as 0.95. Then train SGD classifier on reduced dataset and prints its accuracy and confusion matrix. **10**

UNIT - V

7 a) Explain the following with suitable examples: **10**

- DBSCAN
- Image Segmentation

b) Design the steps involved in clustering the customers based on the similarity using necessary SKlearn libraries, considering using Country clusters.csv. **10**

Country	Latitude	Longitude	Language
USA	44.97	-103.77	English
Canada	62.4	-96.8	English
France	46.75	2.4	French
UK	54.01	-2.53	English
Germany	51.15	10.4	German
Australia	-25.45	133.11	English

******