

PUSASQF602
PREDICTIVE ANALYTICS & MACHINE LEARNING

Time: 2 Hours
Total Marks: 60

Note:

- The candidate has the option to either question 3A or question 3B. Rest all questions are mandatory
- Numbers to the right indicate full marks
- The candidates will be provided with the formula sheet and graphs (if required) for the examination
- Use of approved scientific calculator is allowed

Q1. Answer the following

15 Marks

A.

5 Marks

Perform the following operations mentioned below on the diamonds dataset.

- i. Read the data "Youtuber.csv" using Pandas (1)
- ii. Generate a bar plot of Top 7 Youtube Channels by subscribers
The graph should have titles as mentioned below (2)
 - Title: Top 5 YouTube Channels by Subscribers
 - X Axis Title: Channel Name
 - Y Axis Title: Subscribers (in millions)
- iii. Generate a plot for Distribution of Channels by Country
The graph should have titles as mentioned below (2)
 - Title: Distribution of Channels by Country
 - X Axis Title: Country
 - Y Axis Title: Number of Channels

B.

5 Marks

Load the dataset FIFA19.csv

- i. Filter the data to include only the 'Name', 'Age', 'Nationality', 'Club', 'Value', 'Wage', and 'Overall' columns (1)
- ii. Calculate average "Wage" for each Nationality (1)
- iii. Derive any 2 insights from the data (3)

C.

5 Marks.

i. Run a logistic regression in the below given dataframe

```
df = pd.DataFrame({  
    'Cust_ID': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15],  
    'Salary': [1000, 1100, 10000, 1000, 11000, 1110, 21000,  
30000, 2100, 33000, 21000, 21000, 50000, 21000, 45000],  
    'EMI': [0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1]  
})
```

The data frame consists of 6 employees along with their monthly salaries to check their eligibility for No Cost EMI

Cust_ID: Customer ID for the inquiry

Salary: Customer's monthly take home salary

EMI: Checks eligibility for the EMI

ii. Predict whether the customer is EMI worthy or not (2)

iii. Provide the confusion matrix & score (2)

Q2. Answer the following

15 Marks

A.

5 Marks

Generate a random dataset using the below code:

i. `X, y_true = make_blobs(n_samples=300, centers=4, cluster_std=0.60, random_state=0)` (1)

ii. Plot the dataset. (2)

iii. Apply K Means clustering with suitable number of clusters (2)

B.

5 Marks

Apply Principal Component Analysis on "diamonds.csv" to derive 4 principal components.

C.

5 Marks

Load the covid_19_india dataset in python and perform the below mentioned steps

i. Provide the summarised view of "Cured", "Deaths", "Confirmed" cases per state (3)

ii. Show no. of covid cases with respect to YYYYMM(Year-Month) on x-axis (2)

Q3.

A.

30 Marks

Predict “churn” using the “Bank Customer Churn Prediction.csv”.

- i. Load the dataset (1)
- ii. Get the insights & Correlation for each column vs the output column (5)
- iii. Do the outlier treatment & Null imputation if required. (2)
- iv. Shortlist the most important features for predicting the “churn” (3)
- v. Split the data into features and target (2)
- vi. Perform train test split with a ratio 20% (2)
- vii. Define any 3 classifier models & Train the model on train dataset and predict the model on test dataset (5)
- viii. Calculate the accuracy of the model (2)
- ix. Generate the classification report (2)
- x. Generate the confusion matrix (2)
- xi. Which model is the most suitable one in predicting the output column (4)

OR

Q3.

30 Marks

B.

Predict “Car Purchase Amount” using the “Car_Purchasing_Data.csv”.

- i. Load the dataset (1)
- ii. Get the insights & Correlation for each column vs the output column (5)
- iii. Do the outlier treatment & Null imputation if required. (2)
- iv. Shortlist the most important features for predicting the “Car Purchase amount” (5)
- v. Split the data into features and target (2)
- vi. Perform train test split with a ratio 20% (2)
- vii. Define any 3 regression models & Train the model on train dataset and predict the model on test dataset (5)
- viii. Calculate the accuracy of the model (2)
- ix. Which model is the most suitable one in predicting the output column (6)