2023

COMPUTER SCIENCE — HONOURS — PRACTICAL

Paper : DSE-A-2P

Full Marks : 30

The figures in the margin indicate full marks.

Set - I

Marks Distribution

Source Code : 15
Output : 05
Sessional : 04
Viva voce : 06

Viva voce : At least 3 questions may be asked to each student.

Softwares to be used : Python IDLE/Jupyter Notebook.

All relevant packages should be installed in the computer by the examination centre.

Answer *any one* question.

1. Perform the following actions on the IRIS Dataset :

   (a) Read the csv file into a Pandas Dataframe.

   (b) Display the first 15 records present in the dataset.

   (c) Split the dataset into training set and testing set. The split should be 90% Training set and 10% Testing set.

   (d) Perform classification on the dataset using Logistic Regression. Print the number of features, value of the intercept and the accuracy obtained.   5+3+2+5

2. Perform the following actions on the IRIS Dataset :

   (a) Read the csv file into a Pandas Dataframe.

   (b) Display the details of the columns present in the dataset. The details should include the name and type of the data present in the dataset.

   (c) Split the dataset into training set and testing set. The split should be 80% Training set and 20% Testing set.

   (d) Perform classification on the dataset using Multinomial Naïve-Bayes' Algorithm. Print the accuracy obtained.   5+3+2+5

3. Perform the following actions on the IRIS Dataset :

   (a) Read the csv file into a Pandas Dataframe.

   (b) Display the count of the number of records that belong to each class.

   (c) Split the dataset into training set and testing set. The split should be 60% Training set and 40% Testing set.

   (d) Perform classification on the dataset using k-NN (k Nearest Neighbors). Take the value of k as 3. Print the accuracy.   5+3+2+5

4. Consider the following points :

X	51	51	52	53	52	53	52	11	11	12	13	12	13	12
Y	31	32	31	32	33	33	32	41	42	41	42	43	43	42

   (a) Represent the above points in a csv file.

   (b) Convert the dictionary into a pandas dataframe.

   (c) Plot the points.

   (d) Perform Clustering using the agglomerative clustering algorithm. Print the labels (cluster number) given to each point by the clustering algorithm.   5+3+2+5

5. Perform the following actions on the Titanic Dataset:

   (a) Read the csv file into a Pandas Dataframe.

   (b) Remove the rows which contains empty cells under any of the columns.

   (c) Split the dataset into training set and testing set. The split should be 80% Training set and 20% Testing set.

   (d) Perform classification on the dataset using Logistic Regression. Print the number of features, value of the intercept and the accuracy obtained.   5+3+2+5

6. Perform the following actions on the Titanic Dataset:

   (a) Read the csv file into a Pandas Dataframe.

   (b) Display the details of the columns present in the dataset. The details should include the name and type of the data present in the dataset.

   (c) Split the dataset into training set and testing set. The split should be 80% Training set and 20% Testing set.

   (d) Perform classification on the dataset using Multinomial Naïve-Bayes' Algorithm. Print the accuracy obtained.   5+3+2+5

7. Perform the following actions on the Titanic Dataset:

   (a) Read the csv file into a Pandas Dataframe.

   (b) Display the count of the number of records that belong to each class.

   (c) Split the dataset into training set and testing set. The split should be 60% Training set and 40% Testing set.

   (d) Perform classification on the dataset using k-NN (k Nearest Neighbors). Take the value of k as 3. Print the accuracy.   5+3+2+5

8. Perform the following actions on the Titanic Dataset :

   (a) Read the csv file into a Pandas Dataframe.

   (b) Display the last 10 records present in the dataset.

   (c) Remove the rows which contains empty cells under any of the columns.

   (d) Split the dataset into training set and testing set. The split should be 70% Training set and 30% Testing set.

   (e) Perform classification on the dataset using Gaussian Naïve Buyes' Algorithm. Print the accuracy obtained.   5+3+2+2+3

2023

COMPUTER SCIENCE — HONOURS — PRACTICAL

Paper : DSE-A-2P

Full Marks : 30

The figures in the margin indicate full marks.

Set - II

Marks Distribution

Source Code	:	15
Output	:	05
Sessional	:	04
Viva voce	:	06

Viva voce : At least 3 questions may be asked to each student.

Softwares to be used : Python IDLE/Jupyter Notebook.

All relevant packages should be installed in the computer by the examination centre.

Answer *any one* question.

1. Demostrate the following data preprocessing tasks on a given dataset :
   (a) Dealing with categorical data.
   (b) Scaling the features.
   (c) Splitting dataset into Training and Testing sets.   5+5+5

2. Perform the following actions on the IRIS Dataset :
   (a) Read the csv file into a Pandas Dataframe.
   (b) Display the mean sepal length.
   (c) Split the dataset into training set and testing set. The split should be 92% Training set and 8% Testing set.
   (d) Perform classification on the dataset using Multilayer Perceptron Neural Network Classifier. Print the accuracy obtained.   5+3+2+5

3. Consider the following points:

X	1	2	2	3	2	21	23	21	21	12	13	12	13	12
Y	1	1	2	2	3	23	24	21	22	11	12	13	13	12

(a) Represent the above points as a dictionary. Take "X" and "Y" as keys while the values will be lists containing all the X-coordinates and Y-coordinates respectively.

(b) Plot the points.

(c) Convert the dictionary into a pandas dataframe.

(d) Perform Clustering using the k-means algorithm. Take the value of k as 3. Print the cluster centres.

5+3+2+5

4. Consider the following points:

X	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Y	1	4	6	8	10	12	14	16	18	20	22	24	26	28

(a) Create 2 Numpy arrays X (containing all the X-coordinates) and Y (containing all the Y-coordinates).

(b) Perform regression using the linear regression algorithm.

(c) Print the values of the slope and y-intercept.

(d) Predict the Y-coordinate values for X = [25, 26, 27, 28].

5+3+2+5

5. Perform the following actions on the Titanic Dataset:

(a) Read the csv file into a Pandas Dataframe.

(b) Remove the rows which contains empty cells under any of the columns.

(c) Split the dataset into training set and testing set. The split should be 79% Training set and 21% Testing set.

(d) Perform classification on the dataset using Gaussian Naïve-Bayes' Algorithm. Print the accuracy obtained.

5+3+2+5

6. Perform the following actions on the Titanic Dataset :

(a) Read the csv file into a Pandas Dataframe.

(b) Display the first 10 rows.

(c) Split the dataset into training set and testing set. The split should be 91% Training set and 9% Testing set.

(d) Perform classification on the dataset using Multilayer Perceptron Neural Network Classifier. Print the accuracy obtained.

5+3+2+5

7. Consider the following points:

X	2	2	3	3	4	4	1	11	11	12	13	12	13	12
Y	1	2	1	0	2	1	0	0	1	2	1	1	0	0

(a) Represent the above points in a csv file.

(b) Convert the dictionary into a pandas dataframe.

(c) Plot the points.

(d) Perform Clustering using the agglomerative clustering algorithm. Print the labels (cluster number) given to each point by the clustering algorithm.

5+3+2+5

8. Perform the following actions on IRIS Dataset :

(a) Read the csv file into a Pandas dataset.

(b) Display first 20 records.

(c) Replace missing values with relevant values.

(d) Use bar chart for plotting number of missing values for each attribute.

5+3+2+5

_____

2023

COMPUTER SCIENCE — HONOURS — PRACTICAL

Paper : DSE-A-2P

(Data Mining Laboratory)

Full Marks : 30

Set - 3

Marks Distribution

Source Code	:	15
Output	:	05
Sessional	:	04
Viva voce	:	06

Answer *any one* question.

1. Write a program to perform Data Transformation Technique using Normalization by decimal scaling.

2. Import 'Titanic' dataset. Apply DBSCAN, Agglomerative Clustering Algorithm. Evaluate the performance of the algorithm.

3. Import 'student performance' dataset. Predict reading score and writing score as a Linear Regression Problem. (simulate this problem using various values of epoch and batch size).

4. Import provided dataset. Use bar chart for plotting number of missing values for each attribute.

5. Import provided dataset. Find the range of values for numerical variables present in the dataset.

6. Import provided dataset. Replace the string variables using numerical values.

7. Import provided dataset. Find the number of missing values. Replace the missing values with 0.

8. Apply Logistic Regression on the Titanic dataset to predict the probability of survival of the passengers.

_____

2023

COMPUTER SCIENCE — HONOURS — PRACTICAL

Paper : DSE-A-2P

(Data Mining Laboratory)

Full Marks : 30

Set - 4

Marks Distribution

Source Code	:	15
Output	:	05
Sessional	:	04
Viva voce	:	06

Answer *any one* question.

1. Write a program of cluster analysis using simple K-means algorithm.

2. Import provided dataset. Partition the original dataset into training and test dataset with the ratio of 8 : 2.

3. Import Titanic dataset. Predict the chance of survival by Bayesian classification methods.

4. Import any given dataset. Apply pre-processing techniques on it.

5. Import provided dataset. Identify and print categorical and numerical attributes present in the dataset.

6. Write a program to perform data cleaning techniques using smoothing by bin boundaries.

7. Apply Decision Tree classifier to Titanic dataset. Try this classifier model using different criteria and splitter.

8. Write a program to implement K-Nearest Neighbour Classification Technique.

———————