Printed Page:- 05 **Subject Code**:- **ACSAI0617**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VI - THEORY EXAMINATION (2023 - 2024)

Subject: Programming for Data Analytics

Time: 3 Hours **Max. Marks: 100**

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. *This Question paper comprises of* **three Sections -A, B, & C.** *It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.*

2. *Maximum marks for each question are indicated on right -hand side of each question.*

3. *Illustrate your answers with neat sketches wherever necessary.*

4. *Assume suitable data if necessary.*

5. *Preferably, write the answers in sequential order.*

6. *No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.*

SECTION A **20**

1. Attempt all parts:-

1-a. How do you create a Series in Pandas?(CO1) 1

(a) Using pd.Series() function and passing in a Python list or NumPy array

(b) Using pd.DataFrame() function and passing in a Python dictionary or a list of lists

(c) Using pd.Series() function and passing in a Python dictionary

(d) Using pd.DataFrame() function and passing in a Python list or NumPy array

1-b. Which type of plot is used to show the distribution of a dataset? (CO1) 1

(a) Bar plot

(b) Scatter plot

(c) Line plot

(d) Histogram

1-c. CRAN stands for(CO2) 1

(a) Comprehensive R Architecture Network

(b) Comprehensive R Archive Network

(c) Comprehensive R Automatic Network

(d) Complete R Archive Network

1-d. The CSV files are popular because they are. (CO2) 1

(a) capable of storing large amount of data

(b) easier to create

(c) preferred export and import format for databases and spread sheets

(d) All the above

1-e. The $rename operator logically performs a _____ of both the old name and the new name.(CO3) 1

(a) $unset

(b) $set

(c) $Nested

(d) None of the above

1-f. Amongst which of the following is / are the features of MongoDB? (CO3) 1

(a) Authentication

(b) Encryption

(c) Access control

(d) All of the mentioned above

1-g. Which type of neural network is commonly used for image recognition tasks? (CO4) 1

(a) Recurrent Neural Networks (RNNs)

(b) Convolutional Neural Networks (CNNs)

(c) Feedforward Neural Networks (FNNs)

(d) Radial Basis Function Neural Networks (RBFNNs)

1-h. Which type of neural network is commonly used for processing sequential data such as text? (CO4) 1

(a) Convolutional Neural Networks (CNNs)

(b) Recurrent Neural Networks (RNNs)

(c) Radial Basis Function Neural Networks (RBFNNs)

(d) Generative Adversarial Networks (GANs)

1-i. Variational Autoencoders (VAEs) are primarily used for: (CO5) 1

(a) Data augmentation

(b) Image generation

(c) Representation learning

(d) Text classification

1-j. Which activation function is often used in the output layer of a binary classification neural network? (CO5) 1

(a) Sigmoid

(b) Tanh

(c) ReLU

(d) Softmax

2. Attempt all parts:-

2.a. What are the two primary data structures in Pandas? (CO1) 2

2.b. Explain the purpose of the mutate() function in the dplyr package. (CO2) 2

2.c. What are DML and DDL queries in SQL? (CO3) 2

2.d. Define Word Vectors and discuss their importance in natural language processing tasks. (CO4) 2

2.e. What is the encoder in an autoencoder?(CO5) 2

SECTION B **30**

3. Answer any <u>five</u> of the following:-

3-a. Write a Python program to create a Pandas DataFrame from a dictionary of lists, where the keys are the column names and the values are the data for each column.(CO1) 6

3-b. Describe the process of exploratory data analysis (EDA) and its role in uncovering patterns and relationships in data. (CO1) 6

3-c. Explain the purpose of descriptive statistics in data analysis and discuss commonly used measures such as mean, median, mode, variance, and standard deviation. Provide examples of how to calculate and interpret descriptive statistics using R functions. (CO2) 6

3-d. What is RShiny. What are the several advantages of using it? (CO2) 6

3.e. Explain CAP theorem? How is it applicable to NoSQL systems? (CO3) 6

3.f. How does transfer learning work in TensorFlow, and what are some common use cases for it?(CO4) 6

3.g. Describe the architecture and functionality of Variational Autoencoders (VAEs) in detail, including the role of the encoder, decoder, and latent space. (CO5) 6

SECTION C **50**

4. Answer any <u>one</u> of the following:-

4-a. Describe the process of exploratory data analysis (EDA) in detail, including the steps involved and the techniques used to uncover insights and patterns in the data. (CO1) 10

4-b. Create the following NumPy arrays: (CO1) 10

a) A 1-D array called zeros having 10 elements and all the elements are set to zero.

b) A 1-D array called vowels having the elements ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’.

c) A 2-D array called ones having 2 rows and 5 columns and all the elements are set to 1 and dtype as int.

d) Use nested Python lists to create a 2-D array called myarray1 having 3 rows and 3 columns and store the following data:

2.7, -2, -19

0, 3.4, 99.9

10.6, 0, 13

e) A 2-D array called myarray2 using arange() having 3 rows and 5 columns with start value = 4, step size 4 and dtype as float.

5. Answer any <u>one</u> of the following:-

5-a. Discuss the importance of built-in functions in R and how they contribute to the efficiency of data analysis tasks. Provide examples of commonly used built-in functions and explain their applications in data manipulation, statistical analysis, and visualization. (CO2) 10

5-b. Explain the purpose of flexdashboard in R and how it can be used to create interactive dashboards. Discuss features of flexdashboard such as layout customization, interactivity, and embedding of R code chunks. (CO2) 10

6. Answer any <u>one</u> of the following:-

6-a. Explain the output of the following MongoDB queries. (CO3) 10

a) db.contributor.find({$and: [{branch: "CSE"}, {joiningYear: 2018}]}).pretty()

b) db.contributor.find({$nor: [{salary: 3000}, {branch: "ECE"}]}).pretty()

c) db.contributor.find({salary: {$not: {$gt: 2000}}}).pretty()

6-b. Write a python code to create a table Faculty(FID, FNAME, AGE, DEPARTMENT, SALARY, EXPERIENCE) in SQLite3 and execute the following query using Pyhton Code: (CO3) 10

a). Display records of Faculties who has more than 10 year experience and working in CSBS department.

b). Display records of Faculties who are getting salary more than 150K.

c). Display records of Faculties whose age not in (35,56,78).

d). Display records of Faculties who are working in AIML and DS departments

and having experiene more than 10 years.

e). Change the department of faculties to CS who are in IOT department.

7. Answer any <u>one</u> of the following:-

7-a.	Explain the working of RNN model and compare with CNN?(CO4)	10
7-b.	Explore advanced techniques and methodologies for working with text and sequences in TensorFlow, such as recurrent neural networks (RNNs), long short-term memory (LSTM) networks. (CO4)	10

8. Answer any <u>one</u> of the following:-

8-a.	What is Deep Reinforcement Learning, and how does it differ from traditional Reinforcement Learning?(CO5)	10
8-b.	Explain difference between Generative Adversarial Networks (GAN) and Improved GAN. (CO5)	10