Printed Page:- 04 **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. Which of the following is not a method to read data into a DataFrame in Pandas?(CO1) 1

(a) pd.read_csv()
(b) pd.read_excel()
(c) pd.read_table()
(d) pd.read_html()

1-b. Which module is used to work with regular expressions in Python? (CO1) 1

(a) os
(b) sys
(c) re
(d) math

1-c. The dplyr package can be installed from CRAN using _____. (CO2) 1

(a) installall.packages("dplyr")
(b) install.packages("dplyr")
(c) installed.packages("dpl")
(d) library("dplyr")

1-d. Which function is used th see the shape of a data set in R? (CO2) 1

(a) shape()
(b) dim()

(c) both
(d) none of these

1-e. NoSQL stands for(CO3) 1

(a) Not Only SQL
(b) Not a SQL
(c) New SQL
(d) None of the above

1-f. In MongoDB, Sharding is the process of ___ across multiple machines. (CO3) 1

(a) Drop data records
(b) Storing data records
(c) Truncate data records
(d) All of the mentioned above

1-g. Which of the following layers is commonly used in a CNN for feature extraction?(CO4) 1

(a) Dense layer
(b) Pooling layer
(c) Dropout layer
(d) Activation layer

1-h. What does the term "Up and Running with TensorFlow" typically refer to? (CO4) 1

(a) Running a marathon while coding
(b) Getting started quickly with TensorFlow
(c) Turning off TensorFlow
(d) Flying a drone with TensorFlow

1-i. What is the primary goal of Generative Adversarial Networks (GANs)? (CO5) 1

(a) Dimensionality reduction
(b) Data augmentation
(c) Image generation
(d) Text classification

1-j. Which method is used to prevent overfitting in neural networks? (CO5) 1

(a) Dropout
(b) Batch Normalization
(c) L1 Regularization
(d) All of the above

2. Attempt all parts:-

2.a. What is the purpose of the dot (.) character in regular expressions? (CO1) 2

2.b. What is the difference between a vector and a list in R? (CO2) 2

2.c. What is NoSQL, and how does it differ from traditional SQL databases? (CO3) 2

2.d. Describe the process of training Convolutional Neural Networks (CNNs) in 2

TensorFlow. (CO4)

2.e. Briefly explain the purpose of autoencoders in deep learning. (CO5) 2

SECTION-B 30

3. Answer any five of the following:-

3-a. What is sensitivity analysis, and how is it used in decision making?(CO1) 6

3-b. Discuss the role of descriptive statistics in summarizing and understanding the characteristics of a dataset. (CO1) 6

3-c. What are the different data structures in R? Briefly explain about them.(CO2) 6

3-d. What is List in R? Explain creation, accessing and manipulation of List with example in R. (CO2) 6

3.e. Compare the features of MongoDB with RDBMS. (CO3) 6

3.f. Explain the architecture and working principles of Convolutional Neural Networks (CNNs) in TensorFlow. Discuss how CNNs are used for image recognition tasks.(CO4) 6

3.g. Explain the architecture and working principle of deep neural networks (DNNs) in detail. (CO5) 6

SECTION-C 50

4. Answer any one of the following:-

4-a. Explain (a) What is data wrangling, and why is it important in data analysis? (b) How can you clean data in Pandas, and what are some common data cleaning tasks?(CO1) 10

4-b. Describe the steps involved in hypothesis testing in detail, including formulating null and alternative hypotheses, choosing appropriate test statistics, determining significance levels, and interpreting results. (CO1) 10

5. Answer any one of the following:-

5-a. Explain the purpose of the stringr package in R for string manipulation tasks. Discuss common functions in stringr such as str_detect, str_replace, and str_split, and demonstrate their usage with examples.(CO2) 10

5-b. Do as directed. (CO2) 10

a) You have a dataset from which you want to extract a subset. For example, for a data consisting of employee details, you want to create a subset of employees who are above 30 years and who make salary less than 10,000. How will you execute this in R? The salary dataframe be denoted by the variable ‘age’ and the data is contained within the variable ‘employee_data’.

b) What are is.atomic(), is.vector() and is.numeric() functions responsible for?

c) Suppose your data science project requires the installation of the ggplot2 package. How will you install it?

d) Suppose that I want to know the values in c(1, 2, 6, 3, 19) that are not present in c(2, 6, 14, 3, 15). How can you carry this out using built-in function as well as without it?

6. Answer any one of the following:-

6-a.	Explain MongoDB Shell commands with syntax and example.(CO3)	10
6-b.	Differentiate Between (CO3) a) $in and $nin b) $lt and $lte c) $not and $nor d) $eq and $ne	10

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

7-a.	What are some common activation functions used in CNNs?, Explain in details.(CO4)	10
7-b.	Define a CNN architecture using TensorFlow/Keras for image classification. Include convolutional layers, pooling layers, activation functions, and fully connected layers. (CO4)	10

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

8-a.	Explain the working of the GAN model and also write applications.(CO5)	10
8-b.	Autoencoders are known for their role in unsupervised learning and dimensionality reduction, but could you provide a detailed explanation of their architecture, training process, and applications, particularly in reconstructing input data and learning efficient representations? (CO5)	10