Reg. No.															

Question Paper Code	13493

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2025

Sixth Semester

Computer Science and Engineering (IOT)

20CIEL603 - DATA SCIENCE FOR INTERNET OF THINGS

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	*K – Level*	*CO*
1.	Which of the following is a structured data source? (a) Social media posts (b) Sensor logs (c) Relational database tables (d) Audio recordings	*1*	*K1*	*CO1*
2.	What are the main function of an actuator in an IoT system? (a) To collect data from the environment (b) To process raw data (c) To perform actions based on sensor data (d) To store data in the cloud	*1*	*K1*	*CO1*
3.	Indicate which of the following is unsupervised learning problem(s)? (a) Sorting a set of news articles into four categories based on their titles (b) Predicting the type of interaction (positive/negative) between a new drug and a set of human proteins (c) Identifying close-knit communities of people in a social network (d) Forecasting the stock price of a given company based on historical data	*1*	*K1*	*CO2*
4.	Indicate which of the following is/are regression tasks(s)? (a) Predicting whether an email is spam or not spam (b) Predicting the number of new Covid cases in a given time period (c) Predicting the total number of goals a given football team scores in an year (d) Identifying the language used in a given text document	*1*	*K1*	*CO2*
5.	In an interactive map, how can users get more details about a specific region without cluttering the visualization? (a) Using tooltips that appear on hover (b) Increasing the text labels for each region (c) Removing color from the map (d) Replacing the map with a table	*1*	*K1*	*CO3*
6.	What is a key challenge when designing interactive map visualizations? (a) Handling large datasets efficiently (b) Avoiding the use of colors (c) Displaying only one type of data layer (d) Using only static images	*1*	*K1*	*CO3*
7.	If you want to visualize the distribution of multiple numerical variables together, which Seaborn function is most appropriate? (a) sns.boxplot() (b) sns.pairplot() (c) sns.violinplot() (d) sns.countplot()	*1*	*K1*	*CO4*
8.	If df is a Pandas DataFrame, what does df.iloc[2, 3] return? (a) The entire third column (b) The element in the second row and third column (c) The element in the third row and fourth column (d) An error because .iloc[] is only for rows	*1*	*K1*	*CO4*
9.	Consider the following scenario: You have data on the percentage of different types of mobile devices used to access a website. You want to create a visualization that shows the proportion of each device type in a visually appealing way. What type of Matplotlib plot should you use? (a) Scatter plot (b) Pie chart (c) Histogram (d) Line plot	*1*	*K1*	*CO5*

10.	Which summarization technique is most suitable for real-time fraud detection in banking transactions? (a) Aggregating all transactions at the end of the day (b) Using real-time anomaly detection with streaming summaries (c) Computing batch-based summaries for historical trends (d) Using monthly aggregation of account balances	*1*	*K1*	*CO6*

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11.	Differentiate Business Intelligence (BI) and Data Science.	*2*	*K2*	*CO1*
12.	List the primary and secondary data collection methods.	*2*	*K1*	*CO1*
13.	Show how Regression differs from classification.	*2*	*K2*	*CO2*
14.	Define activation function and mention few common activation functions used.	*2*	*K1*	*CO2*
15.	Compare React JS and Angular JS.	*2*	*K2*	*CO3*
16.	How do you design IoT Data using Time series ,maps and graphs?	*2*	*K1*	*CO3*
17.	How to do data aggregation & statistics using PANDAS?	*2*	*K1*	*CO4*
18.	Write a Python Program to create a data type object.	*2*	*K2*	*CO4*
19.	Compare spaCy vs NLTK.	*2*	*K2*	*CO5*
20.	Write a python code to calculate square root and transpose of an array using Numpy.	*2*	*K2*	*CO5*
21.	Why Stream Processing Frameworks for IoT is needed?	*2*	*K1*	*CO6*
22.	Summarize the Challenges of Real-time Processing in IoT.	*2*	*K2*	*CO6*

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23.	a)	Summarize the importance of data pre-processing and also describe the different techniques such as handling missing values, normalization, and outlier detection with examples.	*11*	*K2*	*CO1*
		OR			
	b)	Paraphrase the major steps involved in Exploratory Data Analysis (EDA)? Illustrate with an example.	*11*	*K2*	*CO1*
24.	a)	Describe SVM, kernels in SVM and elaborate any two kernel functions used in SVM.	*11*	*K2*	*CO2*
		OR			
	b)	Describe the usage of transformation kernel for PCA analysis, if A is the covariance matrix. $A = \begin{bmatrix} 3 & 3 \\ 3 & -4 \end{bmatrix}$.	*11*	*K2*	*CO2*
25.	a)	Identify the core principles of interactive data visualization. How do concepts like clarity, accuracy, efficiency, and user control contribute to effective visual communication?	*11*	*K3*	*CO3*
		OR			
	b)	Develop the steps involved in creating a basic bar chart using D3.js. Discuss how data binding, scales, axes, and SVG elements are used to render the chart. Include a simple code example.	*11*	*K3*	*CO3*
26.	a)	Apply the knowledge of NumPy to perform operations like creating arrays, accessing elements, slicing, and performing arithmetic operations.	*11*	*K3*	*CO4*
		OR			

b) Apply the use of the Matplotlib library for data visualization in Python. Describe the following concepts with appropriate examples: *11 K3 CO4*

(i) Installing and Importing Matplotlib
(ii) Creating Basic Plots with Matplotlib
(iii) Common Plot Types (line plot, bar chart, histogram, etc.)
(iv) Pie Charts in Matplotlib
(v) Use of Markers in Plots

27. a) Experiment the use of the spaCy library for Natural Language Processing (NLP) in Python. Explain the key features of spaCy and demonstrate the following NLP tasks with suitable Python examples *11 K3 CO5*

(i) Loading a spaCy Language Model
(ii) Tokenization
(iii) Part-of-Speech (POS) Tagging
(iv) Named Entity Recognition (NER)

OR

b) Model the use of the Matplotlib library for data visualization in Python. *11 K3 CO5*

28. a) Make use of the concept of Real-Time Data Processing in the context of IoT. Discuss the importance and benefits of using real-time processing in IoT applications. Also, list and describe the key characteristics of real-time data processing in IoT systems. *11 K3 CO6*

OR

b) Develop the role and features of Apache-based stream processing frameworks used in IoT applications. Illustrate how these frameworks help in handling real-time data streams in IoT environments. *11 K3 CO6*