Roll. No:

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech.

SEM: V - THEORY EXAMINATION (2023 - 2024)

Subject: 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 uses data on some object to predict values for other   1
object? [CO1]

        (a) Inferential

        (b) Exploratory

        (c) Predictive

        (d) None of the above

1-b.   Structured data conforms to a data model or scheme and is often stored in   1
tabular form? [CO1]

        (a) TRUE

        (b) FALSE

1-c.   Pictorial representation of data using symbols is known as? [CO2]   1

        (a) Bar Graph

        (b) Pictograph

(c) Pie chart

(d) None of the above

1-d. A teacher notes down the weight of each student in the class. What level of measurement is being used in this case? [CO2]   1

(a) Ordinal

(b) Ratio

(c) Nominal

(d) Interval

1-e. Amongst which of the following step is performed by data scientist after acquiring the data? [CO3]   1

(a) Deletion

(b) Data Replication

(c) Data Integration

(d) Data Cleansing

1-f. I can just impute the mean for any missing data. It won't affect results, and improves power? [CO3]   1

(a) TRUE

(b) FALSE

1-g. Data Wrangling tools includes? [CO4]   1

(a) Spreadsheets

(b) OpenRefine

(c) Data Wrangler

(d) All of the above

1-h. The process of cleaning, organizing and transforming raw data into desired format is known as? [CO4]   1

(a) Data Integration

(b) Data Wrangling

(c) Data Reduction

(d) Data Transformation

1-i. Where can we apply global filters? [CO5]   1

(a) Dashboards

(b) Stories

(c) Sheets

(d) All the above

1-j. Rectangle Boxes used in Tableau for? [CO5]   1

       (a) Dimension

       (b) Measurement

       (c) Tree Map

       (d) All the above

2. Attempt all parts:-

2.a. What is reporting? [CO1]   2

2.b. What does categorical data mean? [CO2]   2

2.c. What is Data Mining? [CO3]   2

2.d. "PCA reduce data dimension", Comment on the following statement? [CO4]   2

2.e. What is the Bar Chart in Tableau? [CO5]   2

SECTION B

30

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

3-a. Explain why is Big Data Important, also state the 5 V's of Big Data computing in detail? [CO1]   6

3-b. Explain the importance of statistics in data science, also state the meaning of P-Value and null value in data perspective? [CO1]   6

3-c. Explain key characteristics of numeric, nominal and categorial data in detail with examples? [CO2]   6

3-d. Explain various advantages of Transactional Data, also explain how can you convert text data to numerical data in data analytics? [CO2]   6

3.e. What causes data inconsistency while preparing for machine learning? Explain with a case study? [CO3]   6

3.f. Describe dimensionality, correlation, eigen vectors in data science perspective in detail? [CO4]   6

3.g. State the components of the dashboard and also explain it's importance in research and decision making with an example? [CO5]   6

SECTION C

50

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

4-a. Explain different Big Data processing software platforms in detail, also explain three application areas of Big Data Processing? [CO1]   10

4-b.   Explain how facebook uses data to revolutionize social networking and   10
         advertising? [CO1]

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

5-a.   What are some of the examples of high dimensional data, explain how it can be   10
         maintained while image processing? [CO2]

5-b.   What are the different types of data sources, also explain Data preprocessing   10
         steps? [CO2]

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

6-a.   What does data cleaning mean? What are the best ways to practice this, explain   10
         steps? [CO3]

6-b.   How many steps are involved in KDD process? Explain in detail with   10
         architecture? [CO3]

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

7-a.   What do understand by Time Series Analysis, explain in detail the Trend and   10
         irregular components of Time Series? [CO4]

7-b.   Explain some techniques we can use to locate outliers in datasets? [CO4]   10

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

8-a.   What is the Difference Between Treemaps and Heat Maps? Explain in detail?   10
         [CO5]

8-b.   Explain six different categories of Tableau Filters? [CO5]   10