JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech III Year I Semester Examinations, September - 2021
DATA ANALYTICS
(Common to CSE, IT)

Time: 3 hours   **Max. Marks: 75**

Answer any five questions
All questions carry equal marks
----

1. Explain about various constraints and influences that will affect data architecture design.   [15]

2. Briefly describe various sources of data like sensors, signals, GPS in data management.
                                                                                                      [15]

3. Discuss various data analytics techniques with examples.   [15]

4. Enumerate your views and observations on different data modelling techniques.   [15]

5. Test for the 'least square estimation' with appropriate case study.   [15]

6. 'Logistic regression is an example of non-linear regression' prove the statement with suitable case study and experiment.   [15]

7. Briefly construct the following decision tree algorithms.
    a) CART
    b) C4.5   [8+7]

8. Interpret on 'pixel-oriented visualization' with example.   [15]

---ooOoo---

Code No: 155AT

R18

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech III Year I Semester Examinations, March - 2021
DATA ANALYTICS
(Common to CSE, IT)

Time: 3 Hours   **Max. Marks: 75**

Answer any five questions
All questions carry equal marks
- - -

1.a) What are hazards? Explain potential sources of hazards in an organization.
  b) What are model-based techniques? Explain.   [8+7]

2.a) Explain about analytics applications to various business domains.
  b) Distinguish between supervised and unsupervised learning.   [8+7]

3.a) Describe measures of forecast accuracy.
  b) Explain about direct visualization and hyperslice.   [8+7]

4.a) Why preprocess the data? Explain in detail.
  b) Explain about variable rationalization with examples.   [8+7]

5.a) Explain in detail about need for business modeling.
  b) What is STL approach? Explain in detail.   [8+7]

6.a) What is wrong with linear regression? Explain logistic regression.
  b) Discuss about dimensional stacking and tree-map.   [8+7]

7.a) Explain about various sources of data in detail.
  b) Discuss in detail about visualizing complex data and relations.   [8+7]

8.   Explain the following:
     a) Apache Spark
     b) Arima
     c) Data visualizations   [5+5+5]

---ooOoo---

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech III Year I Semester Examinations, January/February - 2023
DATA ANALYTICS
(Common to CSE, IT)

Time: 3 Hours   **Max. Marks: 75**

Note: i) Question paper consists of Part A, Part B.
  ii) Part A is compulsory, which carries 25 marks. In Part A, Answer all questions.
  iii) In Part B, Answer any one question from each unit. Each question carries 10 marks
    and may have a, b as sub questions.

PART – A

(25 Marks)

1.a)	Define Data Management.	[2]
b)	Which source of Data is more reliable and Why?	[3]
c)	What are the steps involved in Analysis of Data?	[2]
d)	List out the features of Cloudera Impala.	[3]
e)	Where can be Logistic Regression concepts applied?	[2]
f)	Define Variable Rationalization.	[3]
g)	Mention the applications of Time Series Analysis.	[2]
h)	Name some Software to make Decision Tree.	[3]
i)	What is Chernoff Faces technique?	[2]
j)	What are the advantages of Data Visualization?	[3]

PART – B

(50 Marks)

2.a) What are the constraints and influences that will have an effect on Data Architecture Design? Explain.
  b) Discuss about the Data Reduction as a Data Preprocessing Step.   [5+5]

OR

3.a) What is Secondary Data? Classify the Secondary Data Sources.
  b) Discuss how data quality assessment can depend on the Intended Use of the data? [5+5]

4.a) List and explain the various Analytics techniques.
  b) Explain the various methods to identify the Gaps in the Data and their handling mechanisms.   [4+6]

OR

5.a) Discuss about the importance of Data Modeling in Business.
  b) What are the types of No SQL tools based on Data Models? Explain.   [4+6]

6.a) What is meant by BLUE property? What are the blue properties of OLS method?
  b) Discuss in detail about Multinomial Logistic Regression. [5+5]

OR

7.a) Explain the purpose of the Least Square Estimation in Regression with an example.

  b) Discuss in short about the following Model Fit Statistics:
     i) Hosmer Lemeshow Test   ii) Error Matrix. [5+5]

8.a) What is Overfitting? How to Prevent Overfitting?
  b) Compare and Contrast between ARMA and ARIMA. [5+5]

OR

9.a) Differentiate supervised and unsupervised learning.
  b) Discuss the STL approach for Time Series Decomposition. [5+5]

10.a) Demonstrate Geometric projection visualization techniques.
  b) List out various applications of data visualization. [5+5]

OR

11.a) Write down steps involved in Data Visualization In Tableau.
  b) Explain about tag cloud visualization technique. [5+5]

---ooOoo---

R18

Time: 3 hours   **Max. Marks: 75**

Answer any five questions
All questions carry equal marks
- - -

1. Market researchers have used four experimental designs for generating primary data. Describe them in detail. [15]

2. Explain about data quality and data preprocessing. [15]

3. Demonstrate the various steps involved in data analytics and discuss the tools and environment needed for analytics. [15]

4. Explain how data imputation can be performed. [15]

5.a) Discuss the best unbiased linear estimator property of regression.
  b) Demonstrate ordinary least square estimation. [7+8]

6. Elucidate analytical applications to various business domains. [15]

7.a) What is ETL? List the commercially available ETL tools.
  b) Demonstrate ARIMA for time series data. [7+8]

8. How to perform visualization of the data using a hierarchical partitioning into subspaces? Explain with examples. [15]

---ooOoo---

R18

♦ **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD**
B. Tech III Year I Semester Examinations, August - 2022
DATA ANALYTICS
(Common to CSE, IT)

Time: 3 Hours **Max. Marks: 75**

Answer any five questions
All questions carry equal marks
- - -

1.a) What is data? How to handle large collections of data?
 b) Discuss the various sources of data. [7+8]

2.a) How to identify data quality? What are the quality measures for data?
 b) Explain the need of data preprocessing with illustrations. [7+8]

3.a) Contrast nominal, ordinal and ratio-scaled data.
 b) Explain the applications of data modeling in business. [7+8]

4.a) Illustrate data imputations techniques.
 b) Describe the important features of Cloudera Impala. [7+8]

5.a) Explain ordinary least squares regression with an example.
 b) Illustrate Hosmer–Lemeshow test for goodness of fit of logistic regression. [7+8]

6.a) Discuss the significance of ROC analysis and ROC curve.
 b) Demonstrate variable rationalization in regression. [7+8]

7.a) Explain decision tree induction approach.
 b) How to extract features from time-series data? Explain with an example. [7+8]

8.a) Elucidate visualization of complex data and relations.
 b) Describe parallel coordinates and land scapes for geometric data visualization. [7+8]

---oo0oo---