NARSIMHA REDDY ENGINEERING COLLEGE
UGC-AUTONOMOUS INSTITUTION

An **Autonomous** Institute
NAAC Accreditation 'A' Grade
Accredited by **NBA**
Approved by **AICTE**, Affiliated to **JNTUH**

Department of Artificial Itelligence & Machine Learning

Subject : DATA ANALYTICS[23AM602]

R16

Code No: 138FU

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech IV Year II Semester Examinations, September - 2020
DATA ANALYTICS
(Computer Science and Engineering)

Time: 2 Hours   Max. Marks: 75

Answer any Five Questions
All Questions Carry Equal Marks
- - -

1. Make a comparison of Randomized block design and Latin square design. Quote appropriate examples. [15]

2.a) Explain data preprocessing in data management.
  b) Discuss the process of handling duplicate values in organizational data. [7+8]

3. Explain data imputation and how can repeated imputations enormously improve the quality of estimation. [15]

4.a) Discuss the importance of business modeling.
  b) Compare the techniques for dealing numerical data with categorical data. [7+8]

5. Apply linear regression using the method of least squares to the following data and predict the crop yield for rain fall of 5 cm. [15]

Rain fall(in cms)	10.5	8.8	13.4	12.5	18.8	10.3	7.0	15.6	16
Paddy yield(quintal per acre)	30.3	46.2	58.8	59.0	82.4	49.2	31.9	76.0	78.8

6.a) Explain the advantages of decision trees.
  b) Describe the need of tree pruning in decision trees. [7+8]

7. Discuss in detail the steps involved in ETL process and tools available for this process.[15]

8. Explain the challenges in visualizing complex data and relations and suggest suitable mechanisms to address them. [15]

---ooOoo---

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech IV Year II Semester Examinations, September - 2020
DATA ANALYTICS
(Electronics and Communication Engineering)

Time: 2 Hours

Max. Marks: 75

Answer any Five Questions
All Questions Carry Equal Marks
- - -

1.a) Data set D {10K, 15K,22K, 25K,36K,40K,13K,19K, 88K,94K} represents packages of the students placed in an interview where "*K represents thousand*". Identify the outliers in the data set and analyze its impact in studying the spread of data.
   b) Illustrate techniques of missing values treatment with example. [8+7]

2.a) Demonstrate Missing Imputation methods in detail with examples.
   b) Illustrate Data modeling techniques. [8+7]

3.a) Explain different types of variables used in Regression modeling.
   b) Demonstrate linear regression with suitable example. [8+7]

4.a) Outline major steps of decision tree classification with a suitable example.
   b) What is tree pruning? Illustrate drawback of using separate set of tuples to evaluate pruning. [7+8]

5.a) Apply dimensional stacking and explain how to visualize multivariate data.
   b) Analyze and outline the importance of scales and dimensions in spread sheet visualization. [7+8]

6.a) Demonstrate data preprocessing techniques in detail.
   b) What is data deduplication? Explain deduplication methods. [9+6]

7.a) Qualitative variables are not categorical. Justify with suitable example?
   b) Discuss storage mechanism of unstructured data in distributed computing. [7+8]

8.a) Apply logistic regression to demonstrate binary classification.
   b) What is least square estimate? Illustrate its importance in regression modeling. [7+8]

---ooOoo---