Code No: 138FU

R16

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---