

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---