Code No.: DS403PC

R20 H.T.No. [ ][ ][8][R][ ][ ][ ][ ][ ]

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
II–B.TECH–II–Semester End Examinations (Supply) - July- 2024
INTRODUCTION TO DATA ANALYTICS
(CSD)

[Time: 3 Hours]   [Max. Marks: 70]

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 20 marks. Answer all questions in Part A.
Part B consists of 5 Units. Answer any one full question from each unit. Each question
carries 10 marks and may have a, b, c as sub questions.

PART-A   (20 Marks)

1. a) Define what is meant by 'noise' in a dataset.   [2M]
   b) Define data analytics.   [2M]
   c) Explain the types of databases used in data analytics.   [2M]
   d) List out the types of data.   [2M]
   e) Define regression and write mathematical notation for the linear and non-linear [2M]
      regression models.
   f) What are the applications of linear regression?   [2M]
   g) Write the difference between supervised learning and unsupervised learning.   [2M]
   h) List out the applications of supervised learning models.   [2M]
   i) Define hierarchical visualization and its purpose in data analysis.   [2M]
   j) What is the primary focus of pixel-oriented visualization techniques?   [2M]

PART-B   (50 Marks)

2. Define Data Architecture and explain its components in detail.   [10M]

OR

3. What are outliers in data analysis? Provide a brief explanation with an example and also [10M]
   Describe two methods to handle missing values in a dataset.

4. Explain the steps involved in building a data analytics model for business decision [10M]
   support.

OR

5. Explain the concept of feature selection in data analytics. Why is it important? Provide [10M]
   methods for feature selection with examples.

6. Explain the method of least squares estimation in linear regression. How is the least [10M]
   squares criterion used to derive the regression coefficients?

OR

7. Compare and contrast simple linear regression and multiple linear regression models. [10M]
   Provide examples to illustrate their differences. Also list out the applications of both
   regression models.

8. Define segmentation. How is unsupervised learning used in object segmentation with [10M] examples and applications.

OR

9. List out the models in supervised learning and write in detail about any one of the [10M] supervised learning models.

10. Explain the concept of pixel-oriented visualization. Provide an example of how pixel- [10M] oriented techniques can be used to visualize large dataset.

OR

11. Explain the role of hierarchical visualization techniques in enhancing information [10M] retrieval systems with a case study.

****