



II B.Tech II Sem Regular End Examination, June 2022

Introduction to Data Science

(CSD)

Time: 3 Hours.**Max. Marks: 70**

Note: 1. Question paper consists: Part-A and Part-B.

2. In Part – A, answer all questions which carries 20 marks.

3. 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**(10*2 Marks = 20 Marks)**

- | | | | | |
|-------|---|----|-----|-----|
| 1. a) | What is ETL process? | 2M | C01 | BL1 |
| b) | Define data science technique. | 2M | C01 | BL1 |
| c) | Why Python? | 2M | C02 | BL1 |
| d) | Define Sample and Estimated Mean, Variance and Standard Scores. | 2M | C02 | BL1 |
| e) | What is linear regression? | 2M | C03 | BL1 |
| f) | What is learning? | 2M | C03 | BL1 |
| g) | Define centrality. | 2M | C04 | BL1 |
| h) | What is recommender system? | 2M | C04 | BL1 |
| i) | Define Bi-grams. | 2M | C05 | BL1 |
| j) | What is parallel computing? | 2M | C05 | BL1 |

PART- B**(10*5 Marks = 50 Marks)**

- | | | | | |
|------|---|----|-----|-----|
| 2 a) | Distinguish between structured data and unstructured data. | 5M | C01 | BL2 |
| b) | What are benefits and uses of data science and big data? Explain. | 5M | C01 | BL4 |

OR

- | | | | | |
|------|---|-----|-----|-----|
| 3 | Explain about cleansing, integrating and transforming data in the data science process. | 10M | C01 | BL4 |
| 4 a) | What are fundamental Python libraries for Data Scientists? Explain. | 5M | C02 | BL4 |
| b) | Describe Covariance, and Pearson's and Spearman's Rank Correlation. | 5M | C02 | BL2 |

OR

- | | | | | |
|---|---|-----|-----|-----|
| 5 | Discuss about exploratory data analysis in detail using pandas. | 10M | C02 | BL2 |
|---|---|-----|-----|-----|

- 6 a) Explain about confidence intervals in detail. 5M C03 BL4
b) Discuss about Training, Validation and Test in supervised learning. 5M C03 BL4
- OR**
- 7 Describe TK inter model with suitable examples. 10M C03 BL2
- 8 a) Write about basic definitions in Graphs. 5M C04 BL1
b) How do recommender systems work? Explain. 5M C04 BL4
- OR**
- 9 Define network-X. Explain about networks-X in detail. 10M C04 BL4
- 10 a) Explain the architecture of parallel computing. 5M C05 BL4
b) Describe a Direct View Non-Blocking Proposal of New York Taxi Trips. 5M C05 BL2
- OR**
- 11 Describe statistical natural language processing for Sentiment Analysis with suitable examples. 10M C05 BL2

---oo0oo---