



B.E/B.TECH DEGREE EXAMINATIONS: APRIL/ MAY 2024

(Regulation 2018)

Second Semester

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

U18AII2206: Introduction to Data Science

COURSE OUTCOMES

CO1: Understand the various aspects of data science and the skill sets necessary for a data scientist.

CO2: Explain the concepts of data storage and Big Data.

CO3: Illustrate the different types of processes and tools used in data science.

CO4: Apply the principles of Data Science for analysis using Google Sheets and Excel.

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)
(Answer not more than 40 words)

- | | | | |
|-----|---|-----|-------------------|
| 1. | What are the key skills required for a data scientist? | CO1 | [K ₁] |
| 2. | Define computational thinking in the context of data science. | CO1 | [K ₁] |
| 3. | Differentiate between structured and unstructured data. | CO2 | [K ₂] |
| 4. | What is the significance of data preprocessing? | CO2 | [K ₁] |
| 5. | Describe the concept of data transformation. | CO3 | [K ₂] |
| 6. | How are pivot tables used in data analysis? | CO3 | [K ₁] |
| 7. | What is a probability distribution? | CO3 | [K ₁] |
| 8. | Explain the use of the Z-test in hypothesis testing. | CO4 | [K ₂] |
| 9. | What are the challenges associated with Big Data? | CO2 | [K ₁] |
| 10. | How can Google Sheets be used for data analysis? | CO4 | [K ₁] |

Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11. Scenario:

A retail company wants to analyze its sales data to identify trends and make better business decisions. The data includes sales amounts, dates, product categories, and customer information.

- a) Describe the steps you would take to preprocess this data for analysis. 7 CO1 [K₃]
- b) Explain how you would use pivot tables in Excel to summarize the sales data by product category and date. 7 CO4 [K₂]
- c) What are the benefits of using pivot tables for this type of analysis? 2 CO2 [K₁]

12. **Scenario:**

A healthcare organization is looking to implement a data-driven approach to predict patient readmissions. The dataset includes patient demographics, medical history, treatments, and previous hospital visits.

- a) Outline the process of data cleaning and integration for this dataset. 7 CO2 [K₄]
- b) How would you apply descriptive statistics to understand the distribution of patient ages and treatment outcomes? 7 CO3 [K₃]
- c) Why is data quality important in healthcare analytics? 2 CO2 [K₁]

- 13. a) Explain the concept of hypothesis testing and its importance in data analysis 7 CO3 [K₄]
- b) Explain how you would use the t-test to compare the average sales between two different regions. 7 CO4 [K₂]
- c) What are the assumptions underlying the t-test? 2 CO3 [K₁]

- 14. a) Discuss the challenges of working with unstructured data and the methods to address these challenges. 7 CO2 [K₄]
- b) Explain the process of data discretization and its applications. 7 CO3 [K₅]
- c) How does data reduction help in data analysis? 2 CO2 [K₃]

- 15. a) Describe the different types of data visualization techniques and their applications in data science. 7 CO3 [K₄]
- b) Explain how to create a histogram in Google Sheets and interpret the results. 7 CO4 [K₅]
- c) Why is data visualization important in data science? 2 CO3 [K₂]

- 16. a) Discuss the importance of probability distributions in data science with examples. 7 CO3 [K₂]
- b) Explain the differences between the binomial and normal distributions. 7 CO3 [K₂]
- c) How can random sampling be used to improve data analysis? 2 CO3 [K₃]
