Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 51460

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2024.

Sixth Semester

Civil Engineering

OCS 353 – DATA SCIENCE FUNDAMENTALS

(Common to: Aeronautical Engineering/Aerospace Engineering/Automobile Engineering/Electrical and Electronics Engineering/Electronics and Instrumentation Engineering/Environmental Engineering/Geoinformatics Engineering/Industrial Engineering /Industrial Engineering and Management/Instrumentation and Control Engineering/Manufacturing Engineering/Marine Engineering/Materials Science and Engineering/Mechanical Engineering/Mechanical Engineering (Sandwich)/Mechanical and Automation Engineering/Mechatronics Engineering/Petrochemical Engineering/Production Engineering/Robotics and Automation/Safety and Fire Engineering/Agricultural Engineering/Bio Technology/ Biotechnology and Biochemical Engineering/Chemical Engineering /Chemical and Electrochemical Engineering/Fashion Technology/Food Technology/Handloom and Textile Technology/Petrochemical Technology/Petroleum Engineering/Pharmaceutical Technology/Plastic Technology/Textile Chemistry/Textile Technology)

(Regulations 2021)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. In data science, why data analysis is important?

2. Illustrate the basic statistical descriptions of data.

3. Compare between Python Shell and Jupyter Notebook.

4. Write the process of sorting arrays in NumPy.

5. What is semi-supervised learning, and how does it differ from supervised and unsupervised learning?

6. What are outliers in data analysis? Give example.

7. What are density and contour plots, and when are they used in data visualization?

8. How can text and annotation be added to Matplotlib plots?

9. Name a few techniques for handling large data sets efficiently.

10. What is the importance of data preparation?

PART B — (5 × 13 = 65 marks)

11. (a) Explain in detail the steps involved in the data science process, from defining research goals to presenting findings and building applications.

Or

(b) Compare and contrast data mining and data warehousing, highlighting their respective roles in data science.

12. (a) Discuss the significance of NumPy arrays in data manipulation, explaining universal functions, aggregations, and computation on arrays with relevant examples.

Or

(b) Explain the functionality of pandas in data manipulation for the following.

(i) Data Indexing and selection (4)

(ii) Handling missing data (4)

(iii) Hierarchical Indexing (5)

13. (a) Explain the principles of classification and regression in machine learning, elaborating on their algorithms and real-world use cases.

Or

(b) Explore the concept of clustering in machine learning, discussing different clustering algorithms and their practical implications.

14. (a) Explain the importance of visualizing errors in data and illustrate how it is accomplished in Matplotlib, covering techniques such as error bars and error shading.

Or

(b) Explain the concept of subplots in Matplotlib and how they can be used to create multiple plots within a single figure.

15. (a) Explore various techniques for efficiently managing large data sets, including data partitioning, parallel processing, and distributed computing frameworks, highlighting their advantages and limitations.

Or

(b) Elaborate on the tools and techniques needed for model building and presentation and Automation.

PART C — (1 × 15 = 15 marks)

16. (a) Analyze the case study for building a recommender system used in online shopping.

Or

(b) "Data visualization plays a crucial role in understanding complex datasets and communicating insights effectively". Discuss this statement, elaborating on the techniques and tools available for data visualization, with a focus on Matplotlib and Seaborn.

———————