



CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

Faculty of Technology & Engineering
Computer Science & Engineering Department
Data Science And Analytics [CS442-1]

Marks: 30

Duration: 60 mins.

Section - 1

Answer all the questions.

- 1

What is NoSQL? Differentiate between: SQL and NoSQL.

(5)
- 2

What is MongoDB? Which type of database it is? How can we create collection in MongoDB? Explain with its syntax.

(5)
- 3

Explain the components of SPARK with a diagram.

(5)
- [OR]

4

What is the use of RDD in SPARK? How RDDs can define? Explain with example.

(5)

Section - 2

Answer all the questions.

- 5

You are working on a project that involves analyzing data from multiple sources, including a CSV file, an API, and an SQL database. Discuss the importance of data pre-processing in this context. Describe three specific data cleaning and integration challenges you might encounter while preparing this heterogeneous dataset for analysis. Suggest solutions or techniques to address these challenges.

(5)
- 6

Imagine you have a large dataset with millions of records and dozens of features. Explain the concept of data reduction and its significance in handling such large datasets. Additionally, discuss the process of data discretization and its role in preparing data for visualization. Provide an example of a data visualization technique that would be particularly useful in understanding patterns within the discretized data and why it's effective.

(10)
- [OR]

7

Datasets come in various forms, each with its unique characteristics. Choose a domain of your interest, such as healthcare, finance, or social media, and describe two types of datasets commonly encountered in that domain. Explain the key characteristics and attributes of each dataset type. Furthermore, discuss the relevance and practical applications of these datasets in the chosen domain, highlighting how they contribute to decision-making or insights.

(10)

-----End-----