



TAGORE ENGINEERING COLLEGE

Approved by AICTE, New Delhi | Affiliated to Anna University, Chennai

Rathinamangalam, Vandalur – Kelambakkam Road, Chennai, Tamil Nadu- 600 127, India

Department of Information Technology

ACTIVITY: GROUP DISCUSSION

COURSE NAME: CS3352 Foundations of Data science

Date:19.10.2023

YEAR/SEMESTER: II/III

Faculty: Ms. Gayathri

Topic: Key Pillars of Data Science

In group discussions, students are divided into small groups where they meet face-to-face and freely exchange ideas on a given topic. This collaborative process aims to reach a common decision, enhancing skills that prove valuable when interacting with panel members during placement interviews.

Students discussed on the three foundational areas: Mathematics and Statistics, Programming, and Data Handling.

a. Mathematics and Statistics:

Question: "Why is understanding the mathematics behind algorithms crucial for building robust models?"

- Discussed the role of statistics in understanding and analyzing data. How do statistical methods help interpret data patterns, trends, and relationships?
- Explored probability theory and its importance in predicting outcomes and making decisions under uncertainty.
- Discussed linear algebra and its importance in machine learning algorithms (e.g., how matrices are used to represent datasets and computations).

"Why is understanding the mathematics behind algorithms crucial for building robust models?"

b. Programming:

Question: "How can good programming practices, such as clean code and modularity, enhance the work of a data scientist?"

- Discussed the importance of programming languages like Python, R, and SQL in data manipulation, model building, and automation of tasks.
- Encouraged participants to share their experiences with libraries such as:
 - pandas (for data manipulation and analysis).
 - NumPy (for numerical computing).
 - Matplotlib and Seaborn (for data visualization).
- Discussed real-world use cases where efficient programming has helped solve complex data problems.

c. Data Handling and Management:

Question: "What are the key challenges in data cleaning and preprocessing, and how can they impact the results of data analysis?"

- Discussed the different types of data: structured data (stored in relational databases) and unstructured data (such as images, text, and video).
- Encouraged participants to talk about their experiences with handling large datasets, data cleaning, and preparation (e.g., dealing with missing or inconsistent data).
- Discussed how SQL is used for querying databases and why mastering data retrieval is critical.



Figure:1 Group Discussion