Graphic Era
DEEMED UNIVERSITY

End Term (EVEN) Semester Examination June 2025

Roll no.....

Name of the Course and Semester: M.Tech. CSE & II
Name of the Paper: Data Science and NLP
Paper Code: MCS 251 Maximum Marks: 100

Note:

(i) All the questions are compulsory.
(ii) Answer any two sub questions from a, b and c in each main question.
(iii) Total marks for each question is 20 (twenty).
(iv) Each sub-question carries 10 marks.

Q1. (2X10=20 Marks)

a. Analyze a failed data science project of banking domain and explain which step(s) of the five-step process could have contributed to the failure. CO1

b. Discuss challenges in real-time data collection from IoT devices and how they affect preprocessing pipelines. CO1

c. Explain how data storage strategies differ for structured vs. unstructured data in a cloud-native environment. CO1

Q2. (2X10=20 Marks)

a. Design a machine learning workflow to detect fraudulent transactions. Explain algorithm choice, features, and evaluation. CO2

b. Explore the theoretical limitations of using accuracy as a metric in a multi-class classification problem. CO2

c. In what ways does prior knowledge of probability distributions influence model assumptions in supervised learning? CO2

Q3. (2X10=20 Marks)

a. How can IQR and standard deviation together be used to detect outliers in a mixed-distribution dataset? CO3

b. Discuss regression assumptions and explain what diagnostic plots you would use to validate them. CO3

c. Design a hypothesis test to compare two different machine learning models' performances on the same dataset. CO3

Q4. (2X10=20 Marks)

a. How would you use SQL to create a cohort analysis of users from a SaaS platform? Provide logic and query structure. CO4

b. Discuss performance implications of various JOIN strategies in large-scale analytics databases. CO4

c. Write a SQL script that transforms raw event logs into a user behavior summary using CTEs and window functions. CO4

Q5. (2X10=20 Marks)

a. Build the logic for a domain-specific chatbot (e.g., mental health or banking). Discuss NLP pipeline design and limitations. CO5

b. What are the challenges of applying POS tagging to morphologically rich languages? How can transfer learning help? CO5

c. Explain how transformer models have revolutionized traditional NLP tasks. Compare with RNN-based approaches. CO5