QP Code: D153293	Total Pages: 1	Name:
		Register No.

FIFTH SEMESTER (CUFYUGP) DEGREE EXAMINATION, NOVEMBER 2026

COMPUTER SCIENCE

CSC5CJ301 - DATA MINING

2024 Admission onwards

Maximum Time: 2 Hours | **Maximum Marks: 70**

Section A

All Questions can be answered. Each Question carries 3 marks (Ceiling: 24 Marks)

1	Define data mining. How does it differ from Knowledge Discovery in Databases (KDD)?
2	Distinguish between DBMS and data mining with respect to their objectives.
3	What are the major types of data that can be mined? Give suitable examples.
4	Explain the significance of data cleaning in the data preprocessing process.
5	What is normalization? Explain min-max normalization.
6	Differentiate between frequent itemset and closed itemset.
7	What is the Apriori property? How is it useful in frequent itemset mining?
8	Define Bayes classification and state its basic principle.
9	What is the objective of K-Means clustering?
10	Distinguish between outlier detection and novelty detection.

Section B

All Questions can be answered. Each Question carries 6 marks (Ceiling: 36 Marks)

11	Explain the different types of patterns that can be discovered through data mining.
12	Discuss the major steps involved in data cleaning, with suitable examples of missing and noisy data.
13	Explain entity identification and redundancy analysis problems in data integration.
14	Describe PCA as a data reduction technique.
15	Describe the process of generating association rules from a given set of frequent itemsets.
16	Explain the attribute selection measures used in decision tree construction.
17	Describe the K-Means clustering algorithm and explain its major steps.
18	Explain the working principle of DBSCAN. How does it identify dense regions and noise?

Section C

Answer any ONE. Each Question carries 10 marks (1x10=10 Marks)

19	Explain the KDD process, the role of data mining in KDD, and the major technologies used in data mining.
20	Explain association rule mining using the Apriori algorithm, including frequent itemset generation and rule generation.