

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2026

B.Tech-VI Semester (IT)

COURSE CODE (CREDITS): 18B11CI613 (3)

MAX. MARKS: 15

COURSE NAME: Data Mining

COURSE INSTRUCTORS: Ekta Gandotra

MAX. TIME: 1 Hour

Note: (a) All questions are compulsory.

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

(c) Calculator is allowed.

Q. No.	Question	CO	Marks
Q1.	Explain the architecture of a data warehouse with a neat diagram. Describe the role of each layer and how this architecture supports efficient data analysis.	2	[3]
Q2.	A hospital maintains a data warehouse to analyze its healthcare operations and patient services. The warehouse stores information related to patient visits, doctors, treatments, medicines, time of visit, and hospital location hierarchy (hospital, city, state). The hospital also wants to organize the doctor, medicine, and location information into multiple related tables to reduce data redundancy while supporting analytical queries such as treatment performance by region and medicine usage by department. a. Identify the most appropriate data warehouse schema for this scenario and justify your answer with valid reasons. b. Design the schema structure by identifying the fact table and the related dimension tables.	2	[4]
Q3.	a. Compute the Cosine similarity for the vectors, $x = (0, 1, 0, 1)$, $y = (1, 0, 1, 0)$. b. Calculate the Jaccard similarity between the vectors, $x = (1, 1, 0, 1, 0)$, $y = (1, 0, 1, 1, 0)$. c. Find the correlation coefficient for the vectors $x = (0, 1, 1, 1)$, $y = (1, 1, 1, 0)$. d. Find the Supremum distance between $x = (22, 1, 42, 10)$, and $y = (20, 0, 36, 8)$.	1	[4]
Q4.	The following data represent the number of hours studied by 12 students: 10, 12, 14, 4, 6, 8, 16, 18, 24, 30, 20, 22 Find the five-number summary (Minimum, Q1, Median, Q3, and Maximum) and draw a properly labelled box plot using an appropriate scale.	1	[4]