Name:.....
Roll No:.....

MAR ATHANASIUS COLLEGE OF ENGINEERING (AUTONOMOUS),
KOTHAMANGALAM - 686666
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B.TECH SECOND SERIES EXAMINATION, OCTOBER 2026 (S5 DS,AIML)
B24CD3T01 DATA ANALYTICS

Max. Marks: 40 Duration: 1.5 Hours

PART A

Answer all questions. Each question carries 3 marks.

1. Differentiate between agglomerative and divisive hierarchical clustering (CO3)
2. Explain the concept of clustering and distinguish it from classification. (CO3)
3. Illustrate the concept of One-Hot Encoding with a suitable example. (CO5)
4. Distinguish between stemming and lemmatization with one example for each. (CO5)

PART B

Answer all questions. Each question carries 14 marks.

5. (a) Find the frequent item sets and generate association rules on this. Assume that minimum support threshold (S= 40 Percentage) and minimum confident threshold (c = 70percentage). (CO3)
(7 marks)

ID	ITEMS
T1	E,K,M,N,O,Y
T2	D,E,K,N,O,Y
T3	A,E,K,M
T4	C,K,M,U,Y
T5	C,E,I,K,O

VISION

To mould competent computing professionals, innovators, and leaders guided by ethical values and committed to technological advancement and sustainable societal development.

MISSION

- Provide quality education through strong theoretical foundations, experiential learning, and exposure to emerging computing technologies.
- Cultivate critical thinking, research aptitude, and innovation through problem-based learning, industry interaction, and research initiatives.
- Foster ethical values, leadership, social responsibility, and an aspiration for lifelong learning through professional practices, teamwork, and mentoring.

(b) Explain K-Means clustering algorithm and discuss the steps involved in forming clusters.(CO3) (7 marks)

6. (a) Define Term Frequency and Inverse Document Frequency. Consider a collection of 5 documents in which the term "NLP" occurs 2 times in a document containing 10 terms and appears in 2 documents overall. Calculate the TF, IDF, and TF-IDF of "NLP" using the basic formula. (CO5) (7 marks)

(b) Explain the CBOW model and describe how it predicts a target word from its surrounding context words. Give a suitable example. (CO5) (7 marks)

VISION

To mould competent computing professionals, innovators, and leaders guided by ethical values and committed to technological advancement and sustainable societal development.

MISSION

- Provide quality education through strong theoretical foundations, experiential learning, and exposure to emerging computing technologies.
- Cultivate critical thinking, research aptitude, and innovation through problem-based learning, industry interaction, and research initiatives.
- Foster ethical values, leadership, social responsibility, and an aspiration for lifelong learning through professional practices, teamwork, and mentoring.