VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act 1956)

REG.NO.:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
MID TERM EXAM
SUMMER SEMESTER 2024-2025

SLOT:C1+TC1+C2+TC2

Programme Name & Branch	**: M.TECH (SCOPE)**
Course Code and Course Name	**: CSI4004Text Mining**
Faculty Name(s)	**:Dr.Diviya.M, Dr.A.Sivaranjani**
Class Number(s)	**: VL2024250701092, VL2024250700202**
Date of Examination	**: 09.06.2025**
Exam Duration	**: 90 minutes** **Maximum Marks: 50**

General instruction(s):

Answer All Questions

Q. No	Question	M	CO	BL
1.	What is Named Entity Recognition (NER)? Explain various NER approaches. Given the following sentence, identify and classify the named entities into appropriate categories such as Person, Organization, Location, Date, etc. Sentence: "Apple Inc. was founded by Steve Jobs, Steve Wozniak, and Ronald Wayne on April 1, 1976, in Cupertino, California."	10	CO1	2
2.	D1:The best Italian restaurant enjoy the best pasta D2:American restaurant enjoy the best hamburger D3:Korean restaurant enjoy the best bibimbap D4:th best the best American restaurant Find the coherence across the three documents using TF-IDF	10	CO1	3
3.	A hospital wants to use medical data to diagnose patients. They have various patient characteristics like age, blood pressure, cholesterol levels, etc. (i)Which feature selection method is useful for selecting categorical features (e.g., smoking status, diagnosis history) associated with specific diseases. (ii) Which model can help identify the most important numerical features (e.g., blood pressure, cholesterol levels) that are indicative of diseases. (iii) How would you measure the uncertainties and randomness in your dataset.	10	CO2	2
4.	<table><tr><th>Document</th><th>Term 1 (T1)</th><th>Term 2 (T2)</th></tr><tr><td>D1</td><td>0.8</td><td>0.1</td></tr><tr><td>D2</td><td>0.9</td><td>0.2</td></tr><tr><td>D3</td><td>0.1</td><td>0.9</td></tr><tr><td>D4</td><td>0.2</td><td>0.8</td></tr></table> Use K means clustering to cluster the documents based on the given terms where k=2 with cluster centres as C1 = (0.8, 0.1) and C2 = (0.1, 0.9).	10	CO2	3
5.	Explain in detail the role of Meta algorithms in classification strategies.	10	CO2	2

****