



VIT

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

REG.NO.:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2024-2025

SLOT: C2+TC2

Programme Name & Branch : BTech. CSE, BDS
Course Code and Course Name : BCSE206L, Foundations of Data Science
Faculty Name(s) : Common for all sections
Class Number(s) : Common for all sections
Date of Examination : 18/03/2025
Exam Duration : 90 minutes

Maximum Marks: 50

General instruction(s):

- Answer All Questions
- **Note: Step by step solution is expected.**

C03: Apply advanced tools to work on dimensionality reduction and mathematical operations

C04: Handle various types of data and visualize them using through programming for knowledge representation.

C05: Demonstrate numerous open source data science tools to solve real-world problems through industrial case studies.

Q. No	Question	M	CO	BL
1.	A hospital wants to reduce emergency room (ER) wait times by analysing patient arrival patterns, staff availability, and treatment durations. How can the Data Analytics Lifecycle be applied to optimize ER wait times? Describe each stage and its significance in this context.	10	3	2
2.	Suppose we have the following documents: • Doc1: "machine learning is powerful." • Doc2: "deep learning is a subset of machine learning." • Doc3: "artificial intelligence includes machine learning and deep learning." A user searches for "machine learning applications." Given a query and a set of three documents, remove stop words and apply the Vector Space Model using Term Frequency - Inverted Document Frequency method to order the documents based on cosine similarity.	10	4	3
3.	A university wants to develop an AI-powered chatbot to assist students with course registration. The chatbot must understand student queries like: "I want to register for Data Science. What are the prerequisites? and how can I check the available slots?"	10	4	2



VIT

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

REG.NO.:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2024-2025

SLOT: C2+TC2

	How can Natural Language Processing (NLP) components such as Tokenization, Stopword removal, PoS tagging, Stemming and Lemmatization be applied to improve the chatbot's ability to process and respond to student queries effectively? Provide explanations and significance of each component.	
4.	A university wants to cluster students into different performance groups based on their exam scores out of 10 in Engineering and Science course. We have 5 students with their scores: $S1=(5.3,7.2)$, $S2=(8.5,9.7)$, $S3=(6.1,8.3)$, $S4=(4.4,6.5)$, $S5=(9.9,8.2)$ Construct a distance matrix using the manhattan proximity measure and perform complete linkage hierarchal clustering for the above mentioned two-dimensional data. Show cluster results by drawing a dendrogram.	10
5.	Consider following the Stock Level (Units) and Sales Volume (Units Sold) of four products in a supermarket: $P1=(100,80)$, $P2=(50,30)$, $P3=(200,190)$, $P4=(75,50)$ Compute the Eigen vectors and Eigen values for the given data using the Principal Component Analysis (PCA) Algorithm.	10
