



REG.NO.:

SLOT: C1

Programme Name & Branch : M.Tech (Integrated)-CSE
Course Code and Course Name : CSI3004- DATA SCIENCE PROGRAMMING
Faculty Name(s) : DR.SUBRAMANIVASWAMY V, DR. BERLIN M A, PADMA PRIYA M
Class Number(s) : VL2026270101587, VL2026270101596, VL2026270102383
Date of Examination : 11.08.2026
Exam Duration : 90 minutes
Maximum Marks: 50

General instruction(s):

- Answer All Questions
 - M - Max mark; CO - Course Outcome; BL - Blooms Taxonomy Level (1 - Remember, 2 - Understand, 3 - Apply, 4 - Analyse, 5 - Evaluate, 6 - Create)
- Course Outcomes:
1. Ability to gain basic knowledge on data science
 2. Gain the insights from the data through statistical inferences
 3. Develop suitable models using machine learning techniques and to analyse its performance

Q. No	Question	M	CO	BL
1.	An online shopping company collects customer data from its website, mobile application, transaction records, and customer reviews. Analyse how the different stages of the OSEM framework can be applied to these data sources to identify customer purchasing patterns and generate insights for improving the company's marketing strategy.	10	1	4
2.	Explain the major sources of data available to a Data Scientist. How can Information Commons provide access to useful data from these sources? Discuss with suitable examples.	10	1	2
3.	a) A company has 150 employees, of whom 60 work in the IT department. Among the IT employees, 48 have completed cybersecurity training, while 30% of non-IT employees have also completed the training. Find the probability that an IT employee has completed the training and the probability that a randomly selected employee has completed the training. b) A hospital records 500 patients during a month. Of these, 200 patients are above 60 years of age, and 80 of them were admitted for heart-related problems. Among the remaining patients aged 60 years or below, 40 were admitted for heart-related problems. Calculate the probability of a heart-related admission for a patient above 60 years of age and for a patient aged 60 years or below, and compare the two probabilities.	5	2	3



VIT

Vellore Institute of Technology
(Approved by the University Grants Commission, New Delhi, India)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - I
FALL SEMESTER 2026-2027

REG.NO.:

SLOT: C1

4.	a) An email service receives 25% spam emails and 75% non-spam emails. The word "winner" appears in 80% of spam emails and 3% of non-spam emails. If a randomly selected email contains the word "winner", use Bayes' Rule to determine the probability that the email is actually spam.	5		
	b) A food-processing company wants to estimate the average weight of its packaged products. A random sample of 100 packets has a mean weight of 498 g. The population standard deviation is 15 g. Construct a 95% confidence interval ($Z_{\alpha/2}=1.96$) for the true average weight of the products and interpret the result in the context of quality control.	2	3	

5. A retail company wants to understand whether customer age influences monthly spending. The company collects data from five customers, recording their age and average monthly spending.

Customer	Age (Years)	Monthly Spending (₹ thousands)
C1	22	4.8
C2	27	5.5
C3	32	6.1
C4	37	6.9
C5	42	7.6

10 3 3

The company wants to use Simple Linear Regression to estimate the spending of customers whose ages are outside the observed sample. Apply Linear regression technique to:

- Construct the regression equation for predicting monthly spending based on customer age.
- Calculate the slope and intercept of the regression equation.
- Predict the monthly spending of a customer aged 47 years.
- Predict the monthly spending of a customer aged 52 years.
