CBCS SCHEME

USN | | | | | | | | | | |

BAD702

Seventh Semester B.E./B.Tech. Degree Examination, Dec.2025/Jan.2026
Statistical Machine Learning for Data Science

Time: 3 hrs. Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. M : Marks , L: Bloom's level , C: Course outcomes.

		Module – 1	M	L	C
Q.1	a.	Explain the different measures of central tendency and variability used in EDA. Provide suitable examples.	12	L2	CO1
	b.	Compare and contrast of univariate, bivariate and multivariate data in exploratory data analysis?	08	L2	CO1
		OR			
Q.2	a.	Explain how to explore categorical data. Include both summary statistics and visuals.	12	L2	CO1
	b.	Explain skewness and Kurtosis in data distribution?	08	L2	CO1
		Module – 2			
Q.3	a.	Explain the random sampling and selection bias. How do they affect data analysis? Give examples.	10	L2	CO1
	b.	Describe the concept of confidence intervals. How are they constructed using the normal distribution?	10	L3	CO2
		OR			
Q.4	a.	Explain the Poisson distribution in detail and mention it's application.	10	L2	CO2
	b.	Explain the normal, long-tailed, and student's - t distributions in statistical modeling.	10	L2	CO2
		Module – 3			
Q.5	a.	Describe the concept of P – value in hypothesis testing. Why it is important and how is it interpreted?	10	L2	CO2
	b.	Write the step by step procedure for performing a two – sample t – test?	10	L3	CO3
		OR			
Q.6	a.	What is multiple testing problem? How does it affect statistical inference and how is it corrected?	10	L2	CO3
	b.	What is the role of degrees of freedom in t – tests and other inferential statistics?	10	L2	CO3
		Module – 4			
Q.7	a.	Explain factor (categorical) variables in regression. How are they handled in statistical modeling?	10	L2	CO4
	b.	Describe key regression diagnostics techniques. Explain how they help to improve model quality?	10	L3	CO4

		OR			
Q.8	a.	How do you interpret a regression equation with both continuous and factor variables with example?	10	L2	CO4
	b.	Explain Spline Regression. How is it different from polynomial Regression? Describe its advantages.	10	L3	CO4
		Module – 5			
Q.9	a.	What is the covariance Matric ? Explain it's importance in Discriminant Analysis and how it affect classification boundaries?	12	L2	CO5
	b.	How do we interpret co efficient and odd ratio's in logistic regression? Give an example.	08	L3	CO5
		OR			
Q.10	a.	Explain the precision, Recall, F1 – score. Why are these crucial in imbalanced datasets?	12	L2	CO5
	b.	Describe the ROC curve . How is it used to evaluate classifiers?	08	L2	CO5

* * * * *