	Printed Pages:01

SHARDA UNIVERSITY
Beyond Boundaries

School of Computing Science and Engineering

Department of: Computer Science and Engineering

Assignment-1, Session: 2026-2027

[Programme: B.Tech. CSE] [Semester: VII]

Course Title: Machine Learning Course Code: CSE473		Max Marks: 20
Instructions:	1. All questions are compulsory 2. Assume missing data suitably, if any 3. All questions should be of higher learning.	BTL Level (K)
Unit 1	**CO1: Explain** the fundamental concepts of Machine Learning and underlying stochastic principles.	
Unit 2	**CO2: Compare and differentiate** various Machine Learning models to determine their suitability for solving real-world societal problems.	

Q.No.	**Section A (4×3=12 Marks)**	**Unit**	**COs**	**K**
1	State Mitchell's definition of machine learning. Write the triple ⟨T, P, E⟩ for (a) a spam filter and (b) a handwriting recognizer.	Unit 1	CO1	K1
2	Differentiate between supervised, unsupervised and reinforcement learning on any six points in tabular form. For each, name two algorithms and one real application.	Unit 1	CO1	K2
3	Explain why squared error is not used as the cost function for logistic regression, and write the log-loss (cross-entropy) function.	Unit 2	CO2	K1
4	Explain the different techniques used for handling noisy data. Discuss binning, regression, outlier analysis, and combined human–machine inspection.	Unit 2	CO2	K2
	Section B (2×4=8 Marks)			
5	Explain the four major tasks of data preprocessing. For data cleaning, describe all six standard strategies for handling missing values, and apply mean imputation to the income data 22, 28, ?, 34, 36 (in thousands), stating why the median might be preferred.	Unit 1	CO1	K2
6	For x = (1, 2, 3, 4, 5) and y = (3, 4, 6, 7, 9) with fitted line $\bar{y} = 1.3 + 1.5x$: (a) Compute all five predictions and residuals, and verify that $\sum e_i = 0$. (b) Compute **SSE**, **SST** and R^2. (c) Compute the MSE and interpret the value of R^2	Unit 2	CO2	K2