Department: Information Science and Engineering	**Course Type**: Core
Course Title: Machine Learning Lab	**Course Code:** 18ISL76
L-T-P: 0-0-2	**Credits:**01
Total Contact Hours:26 hours	**Duration of SEE:**3hrs
SEE Marks:50	**CIE Marks:**50

Prerequisite:

- Linear Algebra, Probability & Statistics, Calculus, Data Mining
- Any programming language such as Python, Java

Course Outcomes:

After the course completion, students will be able to:

Cos	Course Outcome Description	Blooms Level
1	Understand the implementation procedures for the machine learning algorithms	L2
2	Design Java/Python programs for various Learning algorithms	L3
3	Apply Machine Learning algorithms to the appropriate data sets	L3
4	Identify and apply Machine Learning algorithms to solve real world problems	L3

Teaching Methodology:

- Black Board Teaching / Power Point Presentation
- Laboratory experiments

Assessment Methods:

- Rubrics for evaluating laboratory experiments for 30 marks
- Two internals, 20 Marks each will be conducted, and average of two internals will be taken.
- Final examination of 50 Marks will be conducted.

Course Outcome to Programme Outcome Mapping:

Cos	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12	PSO 1	PSO 2
CO1	3	3	2		1				2	2				**3**
CO2	3	3	2		1				2	2				**3**
CO3	3	3	2		1				2	2				**3**
CO4	3	3	2		1				2	2				**3**
18ISL76	3	3	2		1				2	2				**3**

Program No.	Title	CO Mapping
Note: • **The programs can be implemented in either Python/Java or any programming language** • **For PART-A, programs are to be developed without using the built-in classes or APIs of Java/Python** • **Data sets can be taken from standard repositories (https://archive.ics.uci.edu/ml/datasets.html) or constructed by the students** • **During Examination Students must execute one program from PART-A and one program from PART-B**		
	PART A	
1.	([115.3, 195.5, 120.5, 110.2, 90.4, 105.6, 110.9, 116.3, 122.3, 125.4]) Use the above array of values and compute the mean, median, mode, Standard deviation, variance min-max normalization and standardization	CO1
2.	Use IRIS data set and perform the PCA on the dataset. Examine the first 2 principal components of X. Create a scatter plot with each of the 150 rows of X projected onto the first two principal components. In other words, the horizontal axis should be first principal component, the vertical axis should be second principal component.	CO1
3.	Given the matrix X whose rows represent different data points, run k-means clustering on this dataset using the Euclidean distance as the distance function. Here k is chosen as 3. The centres of 3 clusters were initialized as μ1 = (6.2, 3.2) (red), μ2 = (6.6, 3.7) (green), μ3 = (6.5, 3.0) (blue).  • What's the centre of the first cluster (red) after one iteration? (Answer in the format of [x1, x2], round your results to three decimal places, same as problems 2 and 3) • What's the centre of the second cluster (green) after two iteration? • What's the centre of the third cluster (blue) when the clustering converges? • How many iterations are required for the clusters to converge?	CO1
4.	Build a Binary Decision Trees using zoo data available at UCI Zoo Data Set. Generate a confusion matrix and print class wise accuracy, precision and recall in your result.	CO2
5.	Use an appropriate 2-dimensional data set and generate scatter plots of its features. Build a correlation matrix and use linear regression to compute the regression parameters. Also compute the Cost, SSE, SSR, SST and R^2.	CO2
6.	Use an appropriate multi-dimensional data set to perform Logistic regression for multi class classification. Illustrate the gradient descent method and compute the regression parameters. Also demonstrate the effect of feature pre-processing like removal of noise, NAN's, Missing value imputation.	CO2
	PART B	
1.	Implement the Naïve Bayesian classifier on COVID data set to predict whether a patient is covid +ve or not. Compute the accuracy, precision recall F1score ROC curve of the classifier, considering 80% training data .Draw the validation curves of the classifier.	CO3

2.	Use RBF, Polynomial and Sigmoid kernel with SVM and compare the performance of the kernels using suitable multiclass data set.	CO3
3.	Build a Random Forest classifier on any readily available disease dataset to predict the correct disease. Compare the performance of the classifier with decision tree.	CO4