DHANALAKSHMI SRINIVASAN ENGINEERING COLLEGE
(AUTONOMOUS)
(Approved by AICTE, New Delhi and Affiliated to Anna University, Chennai)
Re-Accredited with 'A' Grade By NAAC, Accredited by TCS.
Accredited by NBA (AERO, CSE, IT & MECH)
Re-Accredited by NBA (BME, ECE, EEE)
PERAMBALUR - 621212.

LABORATORY COURSE PLAN (2025 – 2026 Even Semester)

LAB COURSE TITLE	**MACHINE LEARNING LABORATORY**			
LAB COURSE CODE	U23CSP42			
LAB COURSE STRUCTURE	LECTURE	TUTORIAL	PRACTICAL	CREDIT
	0	0	3	2
REGULATION	BRANCH	YEAR	SEMESTER	ACADEMIC YEAR
2023	IT	III	VI	2025-2026
COURSE INCHARGE				

SYLLABUS

COURSE OBJECTIVE:

The main learning objective of this course is to prepare the students to:

- To understand the python libraries for data science
- To understand the basic Statistical and Probability measures for data science.
- To learn descriptive analytics on the benchmark datasets.
- To apply correlation and regression analytics on standard datasets.
- To present and interpret data using visualization packages in Python.
- Students will develop the ability to build and assess data

LIST OF EXPERIMENTS

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a.CSV file.

2. For a given set of training data examples stored in a.CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

3. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.

4. Build an Artificial Neural Network by implementing the Back propagation algorithm and test the same using appropriate data sets.

5. Write a program to implement the naïve Bayesian classifier for a sample training dataset stored as a. CSV file .Compute the accuracy of the classifier, considering few test data sets.

6. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.

7. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.

8. Apply EM algorithm to cluster a set of data stored ina.CSV file. Use the same dataset for clustering using k-Means algorithm. Compare the results of the set wo algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.

9. Write a program to implement k -Nearest Neighbour algorithm to classify there is data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.

10. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs

TOTAL: 60 PERIODS

BIBLIOGRAPHY

TEXT/REFERENCE BOOKS:

1. Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw Hill Education.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning, 2nd edition, Springer series in statistics.
3. Ethem Alpaydın, Introduction to machine learning, second edition, MIT press.

HARDWARE:

1. INTEL based desktop PC with min.8GB RAM and 500 GBHDD, 30 17lorhigher TFT Monitor, Keyboard and mouse.
2. Windows10 or higher operating system/Linux Ubuntu 30 20 or higher

SOFTWARE:

1. Python, Numpy, Scipy, Matplotlib, Pandas, statmodels,seaborn, plotly, bokeh
2. Python3.9 or later, Anaconda Distribution, python editors, Jupyter/ PyCharm / equivalent

web link for resource &Virtual lab reference link

W1: https://www.nihlibrary.nih.gov/training/hands-virtual-lab-machine-learning
W2: https://www.vlab.co.in/broad-area-computer-science-and-engineering
W3: https://machinelearningmastery.com/naive-bayes-classifier-scratch-python

EXP. NO.	NAME OF THE EXPERIMENTS	NO. OF PERIODS	CUMULATIVE PERIODS
1.	Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a.CSV file.	8	8
2.	For a given set of training data examples stored in a.CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples	8	16
3.	Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.	4	20
4.	Build an Artificial Neural Network by implementing the Back propagation algorithm and test the same using appropriate data sets	4	24
5.	Write a program to implement the naïve Bayesian classifier for a sample training dataset stored as a.CSV file. Compute the accuracy of the classifier, considering few test data sets	4	28
6.	Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.	4	32
7.	Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.	4	36
8.	Apply EM algorithm to cluster a set of data stored ina.CSV file. Use the same dataset for clustering using k-Means algorithm. Compare the results of the set wo algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the	8	44

	program.		
9.	Write a program to implement k -Nearest Neighbor algorithm to classify there is data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.	8	52
10.	Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs	8	60

COURSE OUTCOME

At the End of the Course, the students will be able to:

CO1: Understand the implementation procedures for the machine learning algorithms.
CO2: Design Java/Python programs for various Learning algorithms.
CO3: Apply appropriate data sets to the Machine Learning algorithms
CO4: Apply Machine Learning algorithms to solve real world problems
CO5: Apply k-Nearest Neighbor algorithm to classify their is data set.
CO6: Apply non-parametric Locally Weighted Regression algorithm

CO-PO Mapping:

CO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO1	2	1	-	-	-	-	-	-	2	1	-	2
CO2	2	1	-	-	-	-	-	-	2	1	-	2
CO3	1	3	2	2	-	-	-	-	1	1	-	2
CO4	1	3	2	2	-	-	-	-	2	1	3	2
CO5	3	2	1	1	-	-	-	-	2	1	-	2
CO6	3	2	1	1	-	-	-	-	-	1	3	2
AVG:	**2.00**	**2.00**	**1.50**	**1.50**	**-**	**-**	**-**	**-**	**1.80**	**1.00**	**3.00**	**2.00**

ADDITIONAL EXPERIMENTS		
EXP. NO.	**NAME OF THE EXPERIMENTS**	**Identified Resource link**
1	Given a dataset. Write a program to compute the Covariance, Correlation between a pair of attributes. Extend the program to compute the Covariance Matrix and Correlation Matrix	https://stackabuse.com/covariance-and-correlation-in-python/
2	Given a set of sample points in N dimensional feature space. Write a program to fit the points with a hyper plane using Linear Regression. Calculate sum of residual error	https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

3	Write a program that provides option to compute different distance measures between two points in the N dimensional feature space. Consider some sample datasets for computing distances among sample points.	https://www.geeksforgeeks.org/program-calculate-distance-two-points/
4	Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.	https://www.goeduhub.com/3127/demonstrate-the-working-the-decision-tree-based-algorithm
5	Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Python ML library classes can be used for this problem.	https://www.vtupulse.com/machine-learning/k-nearest-neighbour-algorithm-in-python/

MODEL LAB DETAILS

BATCH	REGISTER NO.	MODE OF LAB CONDUCT	DATE	TIMING
2023-2027	810423205001-810423205066,301,302,501	OFFLINE		
2023-2027	810423205127-810423205189,306,307	OFFLINE		

VIVA QUESTIONS

1. What's the trade-off between bias and variance?
2. What is the difference between supervised and unsupervised machine learning?
3. How is KNN different from k-means clustering?
4. Define precision and recall.
5. What is Bayes' Theorem? How is it useful in a machine learning context?
6. Why is "Naive" Bayes naive?
7. What's the difference between probability and likelihood?
8. How is a decision tree pruned?
9. What's the F1 score? How would you use it?
10. When should you use classification over regression?
11. How do you ensure you're not over fitting with a model?
12. How would you evaluate a logistic regression model?

13. How do you handle missing or corrupted data in a dataset?
14. How is True Positive Rate and Recall related? Write the equation.
15. What is the difference between supervised and unsupervised machine learning?
16. Comparison between Machine Learning and Big Data
17. What is deep learning?
18. Compare kmeans and EM algorithm
19. Compare find s and candidate elimination algorithm
20. Compare candidate elimination algorithm with decision tree
21. Define numpy
22. Define pandas
23. What is matplotlib? Why it is used?
24. What is tensor flow
25. Define data frame
26. What is hypothesis
27. What is perceptron
28. What is back propagation
29. What is ANN
30. What is feed forward network
31. Define multilayer perceptron
32. Define concept learning.

PREPARED BY
AP/IT

VERIFIED BY
HOD/IT

APPROVED BY

PRINCIPAL