VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY

NAMBUR-522508, ANDHRA PRADESH, INDIA

CourseCode: PC3102 (20CM5T02)	**Subject Title: Machine Learning(CSM)**
Year and Semester: III Year I semester	

Course Objectives:

1. Familiarity with a set of well-known supervised, unsupervised and semi-supervised learning algorithms.
2. Understanding the machine learning model prediction through classification, scoring and ranking using R.
3. Predict objects classification through decision tree building and rule building.
4. Know the importance of features and perform feature engineering
5. Summarizing the data from large tables into smaller set of summary indices through principal component analysis.

Course Outcomes (CO):

1. Explain the differences among the three main styles of Learning: reinforcement learning, supervised, and unsupervised learning.
2. Implement the algorithms for supervised learning and unsupervised learning using R.
3. Determine which of the three learning styles is appropriate to a particular problem domain.
4. Work with real-world data and perform machine learning through data analytics.
5. Characterize the state of the art in learning theory, including its achievements and its challenges.

Unit 1:

Introduction to Statistical Learning: What Is Statistical Learning? , Assessing Model Accuracy.

Linear Regression: Simple Linear Regression, Estimating the Coefficients, Assessing the Accuracy of the Coefficient Estimates, Assessing the Accuracy of the Model.

Unit 2:

Multiple Linear Regression: Estimating the Regression Coefficients, Other Considerations in the Regression Model, Comparison of Linear Regression with K-Nearest Neighbours.

Classification: An Overview of Classification, Why Not Linear Regression? ,Logistic Regression, Generative Models for Classification, A Comparison of Classification Methods.

Unit 3:

Resampling Methods : Cross-Validation, the Bootstrap.

Linear Model Selection and Regularization, Subset Selection, Shrinkage Methods, Dimension Reduction Methods, Considerations in High Dimensions.

Unit 4:

Tree-Based Methods: The Basics of Decision Trees, Regression Trees, Classification Trees, Trees Versus Linear Models, Advantages and Disadvantages of Trees, Bagging, Random Forests, Boosting and Bayesian Additive Regression Trees.

Unit 5:

Support Vector Machines, Maximal Margin Classifier, Support Vector Classifiers, Support Vector Machines.

Unsupervised Learning : The Challenge of Unsupervised Learning, Principal Components Analysis, Missing Values and Matrix Completion, Clustering Methods.

Text Books:

1. Gareth James, et al. An Introduction to Statistical Learning: with Applications in R, Springer. 2nd edition (2021 edition).

Reference Books:

1. Tom Mitchell, "Machine Learning", McGraw Hill, 1997
2. E. Alpaydin, "Introduction to Machine Learning", PHI, 2005.
3. Andrew Ng, Machine learning yearning, https://www.deeplearning.ai/machine-learning-yearning/
4. Hands-on machine learning with R" by Bradley Boehmke & Brandon Greenwell
5. "Machine learning with R, the tidyverse, and mlr" by Hefin I. Rhys

Web Resources:

1. R Programming Crash Course – https://www.youtube.com/watch?v=ZYdXI1GteDE&t=1849s
2. Machine Learning With R | Machine Learning Full Course | Machine Learning Tutorial | Simplilearn - https://www.youtube.com/watch?v=6dEUTmoXz0w
3. Statistics for Data Science | Probability and Statistics | Statistics Tutorial | Ph.D. (Stanford) - https://www.youtube.com/watch?v=Vfo5le26IhY&list=PLlgLmuG_KgbaXMKcISC-fdz7HUn1oKr9i
4. Linear Regression Algorithm | Linear Regression Machine Learning | Linear Regression Full Course - https://www.youtube.com/watch?v=tFi4Y_y-GNM
5. Learning: Support Vector Machines - https://www.youtube.com/watch?v= PwhiWxHK8o

Micro Syllabus of Machine Learning:

Unit 1: **Introduction to Statistical Learning**: What Is Statistical Learning?, Assessing Model Accuracy. **Linear Regression:** Simple Linear Regression, Estimating the Coefficients, Assessing the Accuracy of the Coefficient Estimates, Assessing the Accuracy of the Model.		
Unit No.	**Topic**	**Sub Topic**
I	**Introduction to Statistical Learning**	What Is Statistical Learning, Why Estimate f, How Do We Estimate f
		The Trade-Off Between Prediction Accuracy and Model Interpretability
		Supervised Versus Unsupervised Learning
		Regression Versus Classification Problems
		Assessing Model Accuracy
		Measuring the Quality of Fit
		The Bias-Variance Trade-Off
	Linear Regression	Simple Linear Regression
		Estimating the Coefficients
		Assessing the Accuracy of the CoefficientEstimates
		Assessing the Accuracy of the Model
Unit 2: **Multiple Linear Regression:** Estimating the Regression Coefficients, Other Consid-erations in the Regression Model, Comparison of Linear Regression with K-Nearest Neighbours. **Classification:** An Overview of Classification, Why Not Linear Regression?, Logistic Regression, Generative Models for Classification, A Comparison of Classification Methods.		
II	**Multiple Linear Regression**	Estimating the Regression Coefficients
		Other Considerations in the Regression Model
		Qualitative Predictors
		Extensions of the Linear Model
		Potential Problems
		Comparison of Linear Regression with K-Nearest Neighbours
	Classification	An Overview of Classification
		Why Not Linear Regression?
		Logistic Regression-The Logistic Model
		Estimating the Regression Coefficients
		Making Predictions
		Multiple Logistic Regression
		Multinomial Logistic Regression
		Generative Models for Classification
		Linear Discriminant Analysis for p=1 and p>1
		Quadratic Discriminant Analysis
		Naive Baye's
		A Comparison of Classification Methods- An Analytical Comparison, Empirical Comparison
		Linear Models

Unit 3:

Resampling Methods: Cross-Validation, the Bootstrap.

Linear Model Selection and Regularization, Subset Selection, Shrinkage Methods, Dimension Reduction Methods, Considerations in High Dimensions.

Unit No.	Topic	Sub Topic
III	**Resampling Methods**	Cross Validation-The Validation Set Approach, Leave-One-Out Cross-Validation, k-Fold Cross-Validation
		Bias-Variance Trade-Off for k-Fold Cross-Validation
		Cross-Validation on Classification Problems
		The Bootstrap
	Linear Model Selection and Regularization	Subset Selection-Best Subset Selection, Stepwise Selection
		Choosing the Optimal Model
		Shrinkage Methods
		Ridge Regression
		The Lasso
		Selecting the Tuning Parameter
		Dimension Reduction Methods
		Principal Components Regression
		Partial Least Squares
		Considerations in High Dimensions
		High-Dimensional Data-What Goes Wrong in High Dimensions
		Regression in High Dimensions, Interpreting Results in High Dimensions

Unit 4:

Tree-Based Methods: The Basics of Decision Trees, Regression Trees, Classification Trees, Trees Versus Linear Models, Advantages and Disadvantages of Trees, Bagging, Random Forests, Boosting and Bayesian Additive Regression Trees.

Unit No.	Topic	Sub Topic
IV	**Tree-Based Methods**	The Basics of Decision Trees
		Regression Trees
		Classification Trees
		Trees Versus Linear Models
		Advantages and Disadvantages of Trees
		Bagging
		Random Forests, Boosting
		Bayesian Additive Regression Trees
		Summary of Tree Ensemble Methods

Unit 5:

Support Vector Machines, Maximal Margin Classifier, Support Vector Classifiers, Support Vector Machines.

Unsupervised Learning: The Challenge of Unsupervised Learning, Principal Components Analysis, Missing Values and Matrix Completion, Clustering Methods.

Unit No.	Topic	Sub Topic
V	**Support Vector Machines**	Maximal Margin Classifier-What Is a Hyper plane? Classification Using a Separating Hyper plane, The Maximal Margin Classifier, Construction of the Maximal Margin Classifier
		The Non-separable Case

		Support Vector Classifiers-Overview of the Support Vector Classifier, Details of the Support Vector Classifier
		Support Vector Machines
		Classification with Non-Linear Decision Boundaries
	Unsupervised Learning	The Challenge of Unsupervised Learning
		Principal Components Analysis-What Are Principal Components?, Another Interpretation of Principal Components
		The Proportion of Variance Explained
		More on PCA
		Other Uses for Principal Components
		Missing Values and Matrix Completion
		Clustering Methods-K-Means Clustering, Hierarchical Clustering, Practical Issues in Clustering

****