

III- Year I- Semester	Name of the Course	L	T	P	C
PC3102	Machine Learning	3	0	0	3

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.

Syllabus:

Unit 1-Introduction to Statistical Learning and Linear Regression

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 and Classification

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, Linear Model Selection and Regularization

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

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 and Unsupervised Learning

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.