# FMCE905    DATA SCIENCE (3-1-0)

**Objective:**
- Learn the principles and methods of statistical analysis
- Application of statistical analysis to a range of real-world data sets
- Understand the concept of data analysis using statistics
- Use computational tools on problems of applied nature.

## Module –I (14 Hours)

Linear Methods for Regression and Classification: Overview of supervised learning, Linear regression models and least squares, Multiple regression, Multiple outputs, Subset selection, Ridge regression, Lasso regression, Linear Discriminant Analysis, Logistic regression, Perceptron learning algorithm. Model Assessment and Selection: Bias, Variance and model complexity, Bias-variance trade off, Optimism of the training error rate, Estimate of In-sample prediction error, Effective number of parameters, Bayesian approach and BIC, Cross- validation, Boot strap methods, conditional or expected test error.

**Module  -II :(14 Hours)** Additive Models, Trees ,and Boosting: Generalized additive models**,** Regression and classification trees , Boosting methods-exponential loss and Ada Boost, Numerical Optimization via gradient boosting ,Examples ( Spam data, California housing , New Zealand fish, Demographic data)Neural Networks(NN) , Support Vector Machines(SVM),and K-nearest Neighbour: Fitting neural networks, Back propagation, Issues in training NN, SVM for classification, Reproducing Kernels, SVM for regression, K-nearest –Neighbour classifiers( Image Scene Classification)

## Module -III: (12 Hours)

Unsupervised Learning and Random forests: Association rules, Cluster analysis, Principal Components, Random forests and analysis. Inferential Statistics and Prescriptive analytics: Assessing Performance of a classification Algorithm (t-test,Mc Nemar's test, Paired t-test, paired F-test),Analysis of Variance, Creating data for analytics through designed experiments

**Text Book :**: 1.Trevor Hastie, Robert Tibshirani, Jerome Friedman *, The Elements of Statistical Learning-Data Mining, Inference, and Prediction* ,Second Edition , Springer Verlag,2009.[ chapters:    2,3(3.1-3.4,3.6),4(4.3-4.5),7(excluding    7.8    and    7.9),9(9.1,9.2),(10.1-10.5,10.8,10.10,10.14),11(11.3-11.6),12(12.1-12.3),13.3,14(14.1-14.3.8,14.5.1),15]

**2.** G.James,  D.Witten, T.Hastie, R.Tibshirani-*An introduction to statistical learning with applications                in                R*,Springer,2013.(2.3,3.6.1-3.6.3,4.6.1-4.6.3,5.3,6.6.1,8.3.1,8.3.2,10.4,10.5.1)