B.Tech. CSE (Data Science) III & IV Year

PREDICTIVE ANALYTICS

B.Tech. IV Year I Sem.   **L T P C**
                                                          3 0 0 3

Course Objectives: The course serves to advance and refine expertise on theories, approaches and techniques related to prediction and forecasting.

Course Outcomes

1. Understand prediction-related principles, theories and approaches.

2. Learn model assessment and validation.

3. Understand the basics of predictive techniques and statistical approaches.

4. Analyze supervised and unsupervised algorithms.

UNIT - I

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.

UNIT - II

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.

UNIT - III

Additive Models, Trees, and Boosting: Generalized additive models, Regression and classification trees, Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting, Examples (Spam data, California housing, New Zealand fish, Demographic data).

UNIT - IV

Neural Networks (NN), Support Vector Machines (SVM), and K-nearest Neighbor: Fitting neural networks, Back propagation, Issues in training NN, SVM for classification, Reproducing Kernels, SVM for regression, K-nearest – Neighbour classifiers (Image Scene Classification).

UNIT - V

Unsupervised Learning and Random forests: Association rules, Cluster analysis, Principal Components, Random forests and analysis.

TEXT BOOK:

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, Second Edition, Springer Verlag, 2009.

REFERENCE BOOKS:

1. C.M.Bishop –Pattern Recognition and Machine Learning, Springer, 2006.

2. L. Wasserman-All of statistics.

3. Gareth James. Daniela Witten. Trevor Hastie Robert Tibshirani. An Introduction to Statistical Learning with Applications in R.