INT234:PREDICTIVE ANALYTICS

L:2 T:0 P:2 Credits:3

Course Outcomes: Through this course students should be able to

CO1 :: explain the basics of data preprocessing and its implementation by using R programming Language.

CO2 :: define the basics of classification by using Supervised Learning Algorithms

CO3 :: make use of different Supervised learning techniques to predict numeric values

CO4 :: demonstrate the predictive models by using Neural networks and Support vector machines

CO5 :: classify the data by implementing unsupervised learning algorithms

CO6 :: illustrate the techniques to evaluate the model performance and various methods to improve it

Unit I

DATA PREPROCESSING : Managing data with R, Exploring and understanding data, Exploring the structure of data, Exploring numeric variables, Exploring categorical variables, Exploring relationships between variables

Unit II

SUPERVISED LEARNING: CLASSIFICATION : Lazy learning:Nearest neighbors, Probabilistic Learning: Using Naive Bayes, Divide and Conquer: Decision Trees and Rules

Unit III

SURPERVISED LEARNING : NUMERIC PREDICTION : Polynomial Regression, Ordinary least squares estimation, Correlations, Forecasting Numeric Data, Simple Linear Regression

Unit IV

SUPERVISED LEARNING:DUAL USE : Black Box Methods, Neural Networks, Support Vector Machines

Unit V

UNSUPERVISED LEARNING: CLUSTERING AND PATTERN DETECTION : K-Means Clustering, K-means clustering intuition, K-means random initialization trap, K-means selecting number of clusters, Dataset gathering, Hierarchical Clustering, Association Rules, Finding Patterns, Market Basket Analysis Using Association Rules

Unit VI

MODEL PERFORMANCE : Evaluation Model Performance, Improving Model Performance, Bagging, Boosting, Random forests

List of Practicals / Experiments:

Practical 1: Managing Data with R

- Exploring and understanding the data and loading it into different data structures.

Practical 2: Basics Of Data Preprocessing

- Exploring numeric and categorical variables, and finding relationships between different variables.

Practical 3: Implementation of Lazy and Probabilistic learning algorithms.

- Classification based on Nearest Neighbor and Naïve Bayes.

Practical 4: Implementation of Divide and Conquer Algorithms.

- Classification using Decision Tree and Rules.

Practical 5: Implementation of Regression Algorithms.

- Forecasting using Simple Linear Regression, Polynomial Regression and Multiple Linear Regression Algorithms.

Practical 6: Defining Relationship between Numeric Values.

- Implementation of Ordinary least squares estimation and Correlation algorithms.

Practical 7: Implementation of Dual Supervised Learning Algorithms.

- Black Box Methods, Neural Networks and Support Vector Machines.

Practical 8: Implementation of Clustering Algorithms.

- K-Means Clustering and Hierarchical Clustering.

Practical 9: Implementation of Association Rules.

- Market Basket Analysis Using Association Rules

Practical 10: Model Performance Testing.

- Evaluation Model Performance, Improving Model Performance, Bagging, Boosting, and Random forests

References:

1. PREDICTIVE ANALYTICS: THE POWER TO PREDICT by ERIC SIEGEL, WILEY

2. PYTHON AND R FOR THE MODERN DATA SCIENTIST: THE BEST OF BOTH WORLDS by RICK J. SCAVETTA AND BOYAN ANGELOV, SHROFF/O'REILLY

3. EFFICIENT R PROGRAMMING: A PRACTICAL GUIDE TO SMARTER PROGRAMMING by COLIN GILLESPIE AND ROBIN LOVELACE, SHROFF/O'REILLY

4. APPLIED PREDICTIVE ANALYTICS: PRINCIPLES AND TECHNIQUES FOR THE PROFESSIONAL DATA ANALYST by DEAN ABBOTT, WILEY, 4th Edition, (2012)

5. R IN A NUTSHELL 2E by JOSEPH ADLER, O'REILLY

6. INTRODUCTION TO MACHINE LEARNING WITH R: RIGOROUS MATHEMATICAL ANALYSIS by SCOTT BURGER, SHROFF/O'REILLY