INT234:PREDICTIVE ANALYTICS

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

Course Outcomes: Through this course students should be able to

CO1 :: understand the basics of data preprocessing and exploratory data analysis

CO2 :: apply regression techniques and evaluate regression model performance

CO3 :: apply classification algorithms and assess their performance

CO4 :: analyze clustering algorithms and association rule mining in unsupervised learning

CO5 :: understand and implement foundational neural network models

CO6 :: analyze model performance and apply various methods to enhance predictive accuracy

Unit I

Introduction and Data Preparation : Introduction to Predictive Analytics, machine learning and its types, supervised and unsupervised learning, data preprocessing

Unit II

SUPERVISED LEARNING: REGRESSION : Simple Linear Regression, Multiple Linear Regression, Polynomial Regression, Logistic Regression, Ordinary Least Squares Estimation, Correlations, Evaluate Model Performance: Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared (R^2) Score, Root Mean Squared Error (RMSE)

Unit III

SUPERVISED LEARNING: CLASSIFICATION : Lazy learning:Nearest neighbors, Naïve Bayes, Divide and Conquer: Decision Trees and Rules, Support vector machine, Evaluate Model Performance: Accuracy, Logarithmic Loss, Area Under Curve, Precision, Recall, F1 Score, Confusion Matrix

Unit IV

UNSUPERVISED LEARNING: CLUSTERING AND PATTERN DETECTION : K-Means Clustering, K-means clustering intuition, K-means random initialization trap, K-means selecting number of clusters, Hierarchical Clustering (Agglomerative), Types of Linkages (Single, Complete, Average, Centroid), Association Rules, Finding Patterns, Market Basket Analysis Using Association Rules

Unit V

Dimensionality Reduction and Neural Networks : Dimensionality Reduction, Principal Component Analysis (PCA), Feedforward Neural Networks, Multi-layer Perceptron (MLP)

Unit VI

MODEL PERFORMANCE : Bias-variance trade-off, Ensemble Learning: Basic Concept of Bagging and Boosting, Random forests

List of Practicals / Experiments:

Practical 1:Exploring and handling data

- Exploring datasets, understanding variable types, and loading data into appropriate structures.

Practical 2: Data Preprocessing

- Handling missing data, Exploring numeric and categorical variables, normalization

Practical 3: Regression Modelling

- Implementing Simple Linear Regression, Multiple Linear Regression, Polynomial Regression, and Logistic Regression, Ordinary Least Squares Estimation, Correlations

Practical 4: Classification with Lazy and Probabilistic Learning

- Implementing K-Nearest Neighbors and Naive Bayes.

Practical 5: Decision Trees and SVM Classification

- Building classifiers using Decision Trees and Support Vector Machines.

Practical 6: Clustering Algorithms

- Implementing K-Means clustering and Hierarchical Clustering.

Practical 7: Implementation of Association Rules

- Market Basket Analysis Using Association Rules

Practical 8: Dimensionality Reduction

- Implementing Principal Component Analysis.

Practical 9: Neural Networks

- Implementing Neural Networks.

Practical 10: Model Performance and Ensemble Methods

- Evaluating model performance and implementing Random Forests.

Text Books:

1. INTRODUCTION TO MACHINE LEARNING WITH PYTHON (GREYSCALE INDIAN EDITION) by ANDREAS C. MÜLLER & SARAH GUIDO, SHROFF/O'REILLY

2. DEEP LEARNING FROM SCRATCH: BUILDING WITH PYTHON FROM FIRST PRINCIPLES by SETH WEIDMAN, SHROFF/O'REILLY

References:

1. HANDS-ON MACHINE LEARNING WITH SCIKIT-LEARN, KERAS, AND TENSORFLOW, 3RD EDITION (FULL COLOUR) by AURÉLIEN GÉRON, SHROFF/O'REILLY