DSPC3005 PREDICTIVE ANALYSIS (3-0-0)

Course Objectives:

- To learn, how to develop models to predict categorical and continuous outcomes, using such techniques as neural networks, decision trees, logistic regression, support vector machines and Bayesian network models.
- To know the use of the binary classifier and numeric predictor nodes to automate model selection.
- To advice on when and how to use each model. Also learn how to combine two or more models to improve prediction

Module – I: (08 Hours)

Overview of Predictive Analysis

Introduction to predictive analysis; Predictive Analysis Processing Steps: CRISP-DM; Predictive analysis vs statistics, predictive analysis vs data mining, Application of Modeling in Business, challenges, Defining Data for Predictive Modeling- Defining the Columns as Measures, Defining the Unit of Analysis; Databases & Types of data and variables, Data Modeling Techniques, Missing imputations etc. Need for Business Modelling, Regression — Concepts, Blue property-assumptions-Least Square Estimation, Variable Rationalization, and Model Building etc.

Module – II: (08 Hours)

Data Understanding and Preparation

Introduction, Reading data from various sources, Data visualization, Distributions and summary statistics, Relationships among variables, Extent of Missing Data. Segmentation, Outlier detection, Automated Data Preparation, Combining data files, Aggregate Data, Duplicate Removal, Sampling DATA, Data Caching, Partitioning data, Missing Values.

Module – III: (08 Hours)

Model development & techniques

Data Partitioning, Model selection, Model Development Techniques, Neural networks, Decision trees-The Decision Tree Landscape, Building Decision Trees, Decision Tree Splitting Metrics, Decision Tree Knobs and Options, Other Practical Considerations for Decision Trees, Logistic regression- Interpreting Logistic Regression Models, Other Practical Considerations for Logistic Regression, Model Theory, Discriminant analysis, Support vector machine, Bayesian Networks, Linear Regression, Cox Regression, Association rules.

Module – IV: (08 Hours)

Assessing Predictive Models and Model Ensembles

Batch Approach to Model Assessment, Percent Correct Classification, Rank-Ordered Approach to Model Assessment, Assessing Regression Models; Motivation for Ensembles, Bagging, Boosting, Improvements to Bagging and Boosting- Random Forests, Stochastic Gradient Boosting, Heterogeneous Ensembles, Model Ensembles and Occam's Razor, Interpreting Model Ensembles.

Module – V: (08 Hours)

Model Evaluation and Deployment

Introduction, Model Validation, Rule Induction Using CHAID, Automating Models for Categorical and Continuous targets, Comparing and Combining Models, Evaluation Charts for Model Comparison, Metalevel Modeling, Deploying Model, Assessing Model Performance, Updating a Model.

Course Outcomes:

At the end of this course, students will be able to:

- Understand the process of formulating business objectives and implement predictive models for a various business application.
- Compare the underlying predictive modeling techniques.
- Select appropriate predictive modeling approaches to identify cases to progress with.
- Apply predictive modeling approaches using a suitable package such as SPSS Modeler

Course Outcomes:

After completing this course the student must demonstrate the knowledge and ability to:

CO1: Understand how Predictive analytics can be used in the IT environment.
CO2: Students will grasp the meaning, benefits of Predictive analytics
CO3: Students will understand analyze prediction business capabilities using Time series/Forecasting methods and Extract features

Text Books:

1. "Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst" by Dean Abbott.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, Second Edition, Springer Verlag, 2009.

Reference Books:

1. "Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die" by Eric Siegel.
2. Student's Handbook for Associate Analytics-III.

Weblinks and Video Lectures (E-Resources):

1. https://onlinecourses.swayam2.ac.in/imb20_mg19/preview
2. https://onlinecourses.nptel.ac.in/noc23_ma46/preview