PAPER CODE: MES-OP-303
Applied Predictive Modelling

Course Description

This course describes the predictive modelling process from the perspective of a practitioner rather than a theoretician. It prepares students to gather, describe, and analyse data, and use advanced predictive modelling techniques to make decisions on economics, banking, operations, risk management, finance, marketing, etc. Analysis is done targeting economic and financial decisions in complex systems that involve multiple partners. Broad topics include Data Exploration and Preparation, Prediction and Classification, and Ensemble Methods, Association Rules, Text Analytics, Random Forests

Course Objectives

- To learn to apply concepts of predictive modellingthat enables businesses, in economics, banking and finance, to effectively interpret big data; essential for competition today.
- To learn not only the principles of predictive modelling,but apply them to achieve real, pragmatic solutions.
- To illustrate each technique with hands-on examples, and include in-depth case studies that apply predictive modelling to common economic and business scenarios.

Course Outline

Unit I – Data Exploration and Preparation

Data Cleaning, Handling Missing Data, Identifying Misclassifications, Graphical Methods for Identifying Outliers, Measures of Centre and Spread, Data Transformation, Min–Max Normalization, Z-Score Standardization, Decimal Scaling, Transformations to Achieve Normality, Numerical Methods for Identifying Outliers, Flag Variables, Transforming Categorical Variables into Numerical Variables, Binning Numerical Variables, Reclassifying Categorical Variables, Adding an Index Field, Removing Variables that are not Useful, Variables that Should Probably not be Removed, Removal of Duplicate Records, A Word About ID Fields, Exploring Categorical Variables, Exploring Numeric Variables, Exploring Multivariate Relationships, Selecting Interesting Subsets of the Data for Further Investigation, Using EDA to Uncover Anomalous Fields, Binning Based on Predictive Value, Deriving New Variables: Flag Variables, Deriving New Variables: Numerical Variables, Using EDA to Investigate Correlated Predictor Variables

Unit II – Prediction and Classification

Linear Regression – Estimation, Multicollinearity, Cross-validation, Model adequacy, Parsimony, Penalty-Based Variable Selection in Regression Models with Many Parameters (LASSO);Logistic Regression - Building a Linear Model for Binary Response Data, Interpretation of the Regression Coefficients, Statistical Inference, Classification of New Cases; Binary Classification, Probabilities, and Evaluating Classification Performance;Naive Bayesian Analysis - Model for Predicting a Categorical Response from Mostly Categorical Predictor Variables; The k-Nearest Neighbor Algorithm; Decision Trees - Classification and regression trees (CART), model building, pruning; Chi-Square Automatic Interaction Detection (CHAID); Support Vector Machines (SVM)

Unit III – Ensemble Methods, Association Rules, Text Analytics, Random Forests

Rationale for Using an Ensemble of Classification Models - Bias, Variance, and Noise, Bagging, When to Apply, and not to apply Bagging, Boosting;Affinity Analysis and Market Basket Analysis - Support, Confidence, Frequent Item sets, and the A Priori Property, Generating Frequent Item sets, Generating Association Rules, Extension From Flag Data to General Categorical Data, Information-

Halima Sadia Rizvi
Professor & Head
Deptt. of Economics
Jamia Millia Islamia

Theoretic Approach: Generalized Rule Induction Method;Fundamentals of text data, text mining and sentiment analysis; Random Forests

Textbooks

- *Applied Predictive Modeling*, by Max Kuhn and Kjell Johnson, Publisher:Springer (2013),ISBN 978-1-4614-6848-6
- *Data Mining and Predictive Analytics*, by Daniel T. Larose and Chantal D. larore: Publisher Wiley (2015), ISBN – 978-81-265-5913-8

Reference Books

- *Data Mining and Business Analytics with R*, by Johannes Ledolter; Publisher: Wiley (2013), ISBN-13: 978-1118447147
- *An Introduction to Statistical Learning with Application in R*, by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani; Publisher: Springer (2013); ISBN-13: 978-1461471370
- *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, by Trevor Hastie, Robert Tibshirani and Jerome Friedman; Publisher: Springer (2009), ISBN-13: 978-0387848570

Halima Sadia Rizvi
Professor & Head
Deptt. of Economics
Jamia Millia Islamia
New Delhi-110025