

Subject Code: ID6L001	Name: Data Analytics	L-T-P: 3-0-0	Credit: 3
Prerequisite: None			
<p>Introduction: Sources, modes of availability, inaccuracies, and uses of data.</p> <p>Data Objects and Attributes: Descriptive Statistics; Visualization; and Data Similarity and Dissimilarity.</p> <p>Pre-processing of Data: Cleaning for Missing and Noisy Data; Data Reduction – Discrete Wavelet Transform, Principal Component Analysis, Partial Least Square Method, Attribute Subset Selection; and Data Transformation and Discretization.</p> <p>Inferential Statistics: Probability Density Functions; Inferential Statistics through Hypothesis Tests</p> <p>Business Analytics: Predictive Analysis (Regression and Correlation, Logistic Regression, In-Sample and Out-of-Sample Predictions), Prescriptive Analytics (Optimization and Simulation with Multiple Objectives);</p> <p>Mining Frequent Patterns: Concepts of Support and Confidence; Frequent Itemset Mining Methods; Pattern Evaluation.</p> <p>Classification: Decision Trees – Attribute Selection Measures and Tree Pruning; Bayesian and Rule-based Classification; Model Evaluation and Selection; Cross-Validation; Classification Accuracy; Bayesian Belief Networks; Classification by Backpropagation; and Support Vector Machine.</p> <p>Clustering: Partitioning Methods – k-means Hierarchical Methods and Hierarchical Clustering Using Feature Trees; Probabilistic Hierarchical Clustering; Introduction to Density-, Grid-, and Fuzzy and Probabilistic Model-based Clustering Methods; and Evaluation of Clustering Methods.</p> <p>Machine Learning: Introduction and Concepts: Ridge Regression; Lasso Regression; and <i>k</i>-Nearest Neighbours, Regression and Classification.</p> <p>Supervised Learning with Regression and Classification Techniques: Bias-Variance Dichotomy, Linear and Quadratic Discriminant Analysis, Classification and Regression Trees, Ensemble Methods: Random Forest, Neural Networks, Deep Learning.</p>			

Text/Reference Books:

1. Han, J., M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, Elsevier, Amsterdam. **Textbook.** Year of Publication 2012
2. James, G., D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical learning with Application to R, Springer, New York. Year of Publication 2013
3. Jank, W., Business Analytics for Managers, Springer, New York. Year of Publication 2011
4. Williams, G., Data mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery, Springer, New York. Year of Publication 2011
5. Witten, I. H., E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann. Year of Publication 2011
6. Wolfgang, J., Business Analytics for Managers, Springer. Year of Publication 2011
7. Montgomery, D. C., and G. C. Runger, Applied Statistics and Probability for Engineers. John Wiley & Sons. Year of Publication 2010
8. Samueli G., N. R. Patel, and P. C. Bruce, Data Mining for Business. Intelligence, John Wiley & Sons, New York. Year of Publication 2010
9. Hastie, T., R. T. Jerome, and H. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer. Year of Publication 2009
10. Bishop C., Pattern Recognition and Machine Learning, Springer. Year of Publication 2007
11. Tan, P., M. Steinbach, and V. Kumar, Introduction to Data Mining, Addison-Wesley. Year of Publication 2005