Indian Institute of Information Technology Sri City Chittoor

Course Title: Introduction to Data Analytics (IDA)

Course level: Institute Elective

L-T-P-C : 2-1-0-3

Course Description

This course will provide exposure to theory as well as practical systems and software used in data analytics. The statistical foundations will be covered first, followed by supervised and unsupervised algorithms. Data analytics knowledge could help us understand our world better, and in many contexts enable us to make better decisions.

Course Objective:

The objective of the course is to enable students to graphically interpret data and find meaningful pattern out of it. It also equips students with required statistical tools to model data from various domains and to develop decision support systems.

Course Syllabus:

Units	Topics	Lecture Hours
Unit 1: Data Definitions	Elements, Variables, and Data categorization Levels of Measurement Data management and indexing	4
Unit 2: Descriptive Statistics	Measures of central tendency Measures of location of dispersions, Pre-process the data (Data cleaning, missing data treatment, outliers), Correlation analysis	6
Unit 3: Estimation	Overview of Sampling Theory, Different Sampling Techniques, Important Univariate Distributions (Z, t, Chi Squared, F), Estimation Theory: MLE, MME, Unbiasedness and other properties of a good estimator, RMSE, Standard Error.	6
Unit 4: Parametric and Non-parametric Tests	Statistical hypothesis generation and testing Introduction to Parametric Tests and Non-Parametric Tests Parametric Tests - Z test, t-Test (paired and independent) Non-parametric Tests - Mann Whitney U Test, Sign Test and Wilcoxon signed Rank Test	6
Unit 5: Introduction to Regression and ANOVA	Simple regression Logistic regression ANOVA (Analysis of Variance) - Between Group variability and within Group variability One-way ANOVA Two-way ANOVA	6

Unit 6: Machine Learning techniques	Classification techniques (Bayesian Classifier, Decision trees, SVM) Sensitivity Analysis Similarity measures Clustering Measuring Cluster Goodness Associative Rule Mining	8

Course Assessment or Evaluation:

Minimum attendance required: 75% of the total classes

Examination: 50%
Mid Semester Exam: 20%
End Semester Exam: 30%
Assignment evaluation: 10%
Class Participation: 10%
Scheduled Quizzes: 10%
Project-based evaluation: 20% (10%-10% in two phases)

Resources & References:

1. Douglas C. Montgomery - Design and Analysis of Experiments, 2017, John Wiley & Sons
2. Probability & Statistics for Engineers & Scientists (9th Edn.), Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye, Prentice Hall Inc.
3. The Elements of Statistical Learning, Data Mining, Inference, and Prediction (2ndEdn.), Trevor Hastie Robert Tibshirani Jerome Friedman, Springer, 2014
4. An Introduction to Statistical Learning: with Applications in R, G James, D. Witten, T Hastie, and R. Tibshirani, Springer, 2013
5. Software for Data Analysis: Programming with R (Statistics and Computing), John M. Chambers, Springer
6. Mining Massive Data Sets, A. Rajaraman and J. Ullman, Cambridge University Press, 2012
7. Advances in Complex Data Modeling and Computational Methods in Statistics, Anna Maria Paganoni and Piercesare Secchi, Springer, 2013
8. Data Mining and Analysis, Mohammed J. Zaki, Wagner Meira, Cambridge, 2012
9. Hadoop: The Definitive Guide (2nd Edn.) by Tom White, O'Reilly, 2014
10. MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems, Donald Miner, Adam Shook, O'Reilly, 2014
11. Beginning R: The Statistical Programming Language, Mark Gardener, Wiley, 2013.