

Course Title: <u>DATA ANALYTICS (DA)</u>	L-T-P-C- 3-1-0-4	Institute Elective
Course Description: <p>This course will cover fundamental algorithms and techniques used in Data Analytics. The statistical foundations will be covered first, followed by various machine learning and data mining algorithms. Technological aspects like data management (Hadoop), scalable computation (MapReduce) and visualization will also be covered. In summary, this course will provide exposure to theory as well as practical systems and software used in data analytics.</p>		
Course Objective: <p>After completing this course, the student will learn how to:</p> <ul style="list-style-type: none"> • Find a meaningful pattern in data • Graphically interpret data • Implement the analytic algorithms • Handle large scale analytics projects from various domains • Develop intelligent decision support systems 		
Syllabus: Module I: Data Definitions and Analysis Techniques <ul style="list-style-type: none"> • Elements, Variables, and Data categorization • Levels of Measurement • Data management and indexing Module II: Descriptive Statistics <ul style="list-style-type: none"> • Measures of central tendency • Measures of location of dispersions • Practice and analysis with R Module III: Basic analysis techniques <ul style="list-style-type: none"> • Statistical hypothesis generation and testing • Chi-Square test • t-Test • Analysis of variance • Correlation analysis • Maximum likelihood test 		

Module IV: Data analysis techniques

- Regression analysis
- Classification techniques
- Clustering
- Association rules analysis

Module V: Case studies and projects

- Understanding business scenarios
- Feature engineering and visualization
- Scalable and parallel computing with Hadoop and Map-Reduce
- Sensitivity Analysis

Course Evaluation plan:

Minimum attendance required: 75% of the total classes

Mid-Semester evaluation: 30%

End-Semester evaluation: 30%

Scheduled Quizzes: 10%

Project-based evaluation: 30% (15%-15% in two phases)

Note: Minimum attendance and presence in all evaluations are must (other than some medical or emergency ground. No compensatory test or extended submission).

Resources & References:

1. Probability & Statistics for Engineers & Scientists (9th Edn.), Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye, Prentice Hall Inc.
2. The Elements of Statistical Learning, Data Mining, Inference, and Prediction (2ndEdn.), Trevor Hastie Robert Tibshirani Jerome Friedman, Springer, 2014
3. An Introduction to Statistical Learning: with Applications in R, G James, D. Witten, T Hastie, and R. Tibshirani, Springer, 2013
4. Software for Data Analysis: Programming with R (Statistics and Computing), John M. Chambers, Springer

5. Mining Massive Data Sets, A. Rajaraman and J. Ullman, Cambridge University Press, 2012
6. Advances in Complex Data Modeling and Computational Methods in Statistics, Anna Maria Paganoni and Piercesare Secchi, Springer, 2013
7. Data Mining and Analysis, Mohammed J. Zaki, Wagner Meira, Cambridge, 2012
8. Hadoop: The Definitive Guide (2nd Edn.) by Tom White, O'Reilly, 2014
9. MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems, Donald Miner, Adam Shook, O'Reilly, 2014
10. Beginning R: The Statistical Programming Language, Mark Gardener, Wiley, 2013.