

ELECTIVE II

2CSE50E15: Data Science & Analytics

[3 0 2 3 1]

Learning Outcomes:

After learning the course the students should be able to

- Learn the fundamentals of data analytics and the data science pipeline
- Learn how to scope the resources required for a data science project
- Apply statistical methods, regression techniques, and machine learning algorithms to make sense out of data sets both large and small
- Demonstrate knowledge of statistical data analysis techniques utilized in business decision making.
- Apply principles of Data Science to the analysis of business problems.
- Use data mining software to solve real-world problems.
- Employ cutting edge tools and technologies to analyze Big Data.
- Apply algorithms to build machine intelligence.

SYLLABUS

Unit No.	Topics	Lectures (Hours)
1	Descriptive Statistics Introduction to the course, Descriptive Statistics, Probability Distributions	5
2	Inferential Statistics Inferential Statistics through hypothesis tests, Permutation & Randomization Test	5
3	Regression & ANOVA Regression, ANOVA (Analysis of Variance)	6
4	Machine Learning Introduction and Concepts Differentiating algorithmic and model based frameworks, Regression: Ordinary Least Squares, Ridge Regression, Lasso Regression, K Nearest Neighbours, Regression & Classification	7
5	Supervised Learning with Regression and Classification techniques Bias-Variance Dichotomy, Model Validation Approaches, Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Regression and Classification Trees, Support Vector Machines, Ensemble Methods: Random Forest, Neural Networks, Deep learning	7
6	Unsupervised Learning and Challenges for Big Data Analytics Clustering, Associative Rule Mining, Challenges for big data analytics	5
7	Prescriptive analytics Creating data for analytics through designed experiments, creating data for	6

	analytics through Active learning, creating data for analytics through Reinforcement learning	
8	Visualization Graph Visualization, Data Summaries, Model Checking & Comparison	4

Reference Books:

1. Hastie, Trevor, et al. The elements of statistical learning. Vol. 2. No. 1. New York: springer, 2009.
2. Montgomery, Douglas C., and George C. Runger. Applied statistics and probability for engineers. John Wiley & Sons, 2010
3. Bekkerman et al. Scaling up Machine Learning
4. Tom White “Hadoop: The Definitive Guide” Third Edition, O’reilly Media, 2012.
5. AnandRajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, Cambridge University Press, 2012.
6. Vincent Granville, Developing Analytic Talent: Becoming a Data Scientist, wiley, 2014.
7. Jeffrey Stanton & Robert De Graaf, Introduction To Data Science, Version 2.0, 2013.