

Third Year B.C.A. (Under Science) Semester V or VI

Course Code: BCA-604

Course Title: Data Analytics

**Total Contact Hours: 48 hrs.
(60 Lectures)**

Total Credits: 04

Total Marks: 100

Teaching Scheme: Theory- 05 Lect./ Week

Course Objectives:

1. Able to apply fundamental algorithmic ideas to process data.
2. Learn to apply hypotheses and data into actionable Predictions.

Unit No.	Contents	No. of Lectures
Unit 1	• Introduction to data Science ○ Basics of Data ○ What is Data Science? ○ Data science process ○ Stages in data science project • Basics of Data Analytics • Types of Analytics – Descriptive, Predictive, Prescriptive • Statistical Inference ○ Populations and samples ○ Statistical modeling, ○ Probability ○ Distribution ○ Correlation ○ Regression	10
Unit 2	Introduction to Machine Learning • Basics of Machine Learning • Supervised Machine Learning ▪ K- Nearest-Neighbors, ▪ Naïve Bayes ▪ Decision tree ▪ Support Vector Machines • Unsupervised Machine Learning ▪ Cluster analysis ▪ K means ▪ Association Rule Mining • Apriori algorithms • Regression Analysis ▪ Linear Regression ▪ Nonlinear Regression	25

Unit 3	Data Analytics with Python Programming Numpy □ ○ Arrays ○ Array indexing ○ Datatypes ○ Array math ○ Broadcasting □ SciPy ○ Image operations ○ Distance between points □ Data analysis and manipulation using Pandas package ○ Importing Data , Creating A DataFrame, ○ DataFrame Methods, ○ Indexing DataFrames, Boolean Indexing ○ Indexing Using Labels , Multi-Indexing ○ Merge DataFrames ○ Sorting DataFrames ○ Apply Function ○ Pivot Table, Crosstab - Iterating over rows of a dataframe	15
Unit 4	Data Visualization ▪ Basic principles, ▪ Ideas and tools for data visualization ▪ Graph Visualization, ▪ Data Summaries, ▪ Model Checking & Comparison ▪ Purpose of visualization ▪ Multidimensional visualization ▪ Tree visualization ▪ Graph visualization ▪ Visualization techniques ▪ Understanding analytics output and their usage ▪ Scikit package ▪ matplotlib library ○ Plotting ○ Subplots ○ Images	10

Reference Books:

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