NAME OF THE SUBJECT	L	T	P	C
INTRODUCTION TO DATA SCIENCE	3	0	0	3

Unit – I: Introduction **(10 Hours)**

Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields – Data Security Issues. Data Collection Strategies, Data Categorization: NOIR Topology.

Unit – II: Data Pre-Processing & Exploratory Data Analysis **(12 Hours)**

Descriptive Statistics – Mean Standard Deviation, Skewness and Kurtosis – Box Plots – Pivot Table – Heat Map Correlation Statistics –ANOVA.

Data Pre-Processing Overview – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization.

Unit – III: Model Development **(12 Hours)**

Simple and Multiple Regression – Model Evaluation using Visualization – Residual Plot – Distribution Plot – Polynomial Regression and Pipelines – Measures for In-sample Evaluation – Prediction and Decision Making.

Unit – IV: Model Evaluation **(12 Hours)**

Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Over fitting – Under Fitting and Model Selection – Prediction by using Ridge Regression – Testing Multiple Parameters by using Grid Search.

Text Books:

1. 1. JojoMoolayil, “Smarter Decisions: The Intersection of IoT and Data Science”, PACKT, 2016.
2. 2. Cathy O’Neil and Rachel Schutt , “Doing Data Science”, O'Reilly,2015.
3. 3. David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, EMC 2013
4. 4. Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGIGlobal.