# Course Description

**Course Title: Data Science Course Code: 17B11CI611**

**L-T-P Scheme: 3-1-0 Credit: 4**

**Prerequisite: Students must have already studied the course “Business Analysis Techniques”**

**Course Objectives:**

This course will introduce students to this rapidly growing field and equip them with some of its basic principles and tools as well as its general mindset. Students will learn concepts, techniques and tools they need to deal with various facets of data

Science practice, including data collection and integration, exploratory data analysis, predictive modelling, descriptive modelling, data product creation, evaluation, and effective communication.

**Learning Outcomes:**

At the end of the course students should be able to:

Students will develop relevant programming abilities.

Students will demonstrate proficiency with statistical analysis of data.

Students will develop the ability to build and assess data-based models.

Students will execute statistical analyses with professional statistical software.

Students will demonstrate skill in data management.

Students will apply data science concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively

**Course Content:**

**Unit I:** **Introduction and Data Pre-processing**

Data Science Introduction

Big Data and Data Science

Current landscape of perspectives

**Unit II: Data Analysis and Correlations: Basic Concepts and Methods**

Populations and samples

Statistical modelling, probability distributions, Regression, fitting a model

Dimensionality Reduction: PCA & DWT, Correlation and regression analysis.

Chisquare t and F distributions (definitions only) Confidence interval Single mean and difference known and unknown variances.

**Unit III: Introduction to machine learning and Cluster Analysis : Basic Concept and Methods**

Supervised and unsupervised learning, Training and testing data, Over fitting and under fitting. Distance measures :- Manhattan, Chebbychev, Mahalanobis Distance

Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods,

Grid-Based Methods, Evaluation of Clustering, Clustering High-Dimensional Data,

Clustering Graph and Network Data

**Unit IV: Classification Algorithms**

Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based

Classification, Model Evaluation and Selection, Techniques to Improve Classification

Accuracy, Support Vector Machines, Lazy Learners (or Learning from Your Neighbors)

**Unit V: Introduction to Web Search and Social Media Analytics**

Data Wrangling: APIs and other tools for scrapping the Web

Mining Complex Data Types, Other Methodologies of Data, Mining, Data Mining Applications, Data Mining and Society, Data Mining Trends

Social Media Analytics is the science of analyzing data to convert information to useful knowledge. This knowledge could help us understand our world better and, in many contexts, enable us to make better decisions.

**Evaluation Scheme:**

|  |  |  |
| --- | --- | --- |
| **Exams** | **Marks** | **Coverage** |
| Test-1 | 15 Marks | Based on Unit-1, Unit-2 |
| Test-2 | 25 Marks | Based on Unit-3 & Unit-4 and around 30% from coverage of Test-1 |
| Test-3 | 35 Marks | Based on Unit-5 around 30% from coverage of Test-2 |
| Assignment | 10 Marks |  |
| Tutorials | 5 Marks |  |
| Quiz | 5 Marks |  |
| Attendance | 5 Marks |  |
| **Total** | **100 Marks** |  |

**Text Books:**

1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from the Frontline.

O'Reilly. 2014.

2. Avrim Blum, John Hopcroft and Ravindran Kannan. Foundations of Data Science.

**Refernce Books:**

1. Jure Leskovek, Anand Rajaraman and Je\_rey Ullman. Mining of Massive Datasets. v2.1,
2. Cambridge University Press. 2014.
3. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. ISBN 0262018020. 2013.
4. Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking. ISBN 1449361323. 2013.
5. Trevor Hastie, Robert Tibshirani and Jerome Friedman. Elements of Statistical Learning,
6. Second Edition. ISBN 0387952845. 2009.