# 3. FOUNDATIONS OF DATA SCIENCE

| Semester | Hours/ Week | | | Total hrs | Credit | Max Marks | | |
|---|---|---|---|---|---|---|---|---|
| | L | T | P | | C | CIE | SEE | TOTAL |
| II | 3 | 0 | 0 | 48 | 3 | 40 | 60 | 100 |

## MODULE – 1      Conceptual Introductions      10H

**Descriptive Statistics:** Measures of central tendency – mean, median, mode, hormonic mean and geometric mean.

**Measures of dispersion:** mean deviation from mean, standard deviation and variance.

**Central moments:** Covariance and correlation, rank correlation.

**Sampling distributions:** Hypothesis testing, definition of random variable and probability.

**Probability distributions:** Bernoulli, Binomial, Poisson.

**Continuous probability distributions:** Gaussian, exponential, Chi-square. Definition of Bayes probability. What Is data science, How does data science relate to other fields, Eigen values &Eigen vectors, Sparse matrices.

## MODULE – 2      Python for Data Science      10H

**Python for Data Analysis:** Introduction to Numpy, Numpy Arrays and indexing, Introduction to pandas, Series, Data frames, Missing data, Groupby, Merging Joining and Concatenating, read csv and json, Cleaning Data.

**Python for Data Visualization:** Matplotlib library, Seaborn Distribution, Matrix and Regression Plots, Introduction to SKlearn and Plotly.

## MODULE – 3      Regression      10H

Data Preprocessing in Python, Regression, Simple Linear regression, Multiple Linear Regression, Polynomial regression, Support Vector Regression(SVR), Decision Tree Regression.

## MODULE – 4Supervised Learning -Classification      9H

**Introduction to Supervised Learning:** Logistic Regression, K-Nearest Neighbors(KNN), Support Vector Machine (SVM), NaïveBayes, Decision Tree Classification, Random Forest Classification.

## MODULE – 5Unsupervised Learning -Clustering      9H

**Introduction to Unsupervised Learning:** K-Means Clustering, Hierarchical Clustering, Introduction to Reinforcement Learning, Principal Component Analysis(PCA), Linear Discriminant Analysis(LDA).

**Total Hours : 48Hours**

**Text Book(s):**
1. A Hands On Introduction to Data Science, Cambridge University Press.
2. Principles of Data Science learn the techniques and math you need to start making sense of your data by Sinan Ozdemir.

**Reference Book(s):**
1. Python Data Science Handbook by Jake Vander Plas Released November 2016, O'Reilly Media.
2. JoelGrus, Data Science from Scratch, Oreillymedia, 2015.
3. Linear Algebra and Its Applications, 4thEdition, GilbertStrang.