CS1704 FOUNDATIONS OF DATA SCIENCE + LAB
L T P C
3 0 2 4

COURSE OBJECTIVES:

- CO1: To learn fundamentals of Data Science using Python, to learn about visualization
- CO2: To understand probability distributions and statistical Inferences
- CO3:To be familiar with supervised and unsupervised methods in machine learning
- CO4: To explore the algorithms used for analysing massive data problems

---

INTRODUCTION TO DATA SCIENCE: (9 periods)

Introduction: Need for data science - Benefits and uses - Facets of data - Big data ecosystem -
The data science process: Retrieving data - Cleansing, integrating and transforming data - Data analysis - Build the models - Presenting findings and building applications

---

TOOLBOXES FOR DATA SCIENTISTS: (10 periods)

Introduction to Python - Fundamental Python Libraries for Data Scientists:
Numpy - Scipy - ScikitLearn - Pandas - Matplotlib - IDE - Data Manipulation with Python

---

DESCRIPTIVE STATISTICS: (12 periods)

Introduction - Data Preparation - Exploratory Data Analysis:
Data summarization - Data distribution - Outlier Treatment - Measuring asymmetry - Continuous distribution;
Estimation: Mean - Variance - Sampling - Covariance - Correlation

---

MACHINE LEARNING: (14 periods)

Supervised Learning:
Introduction - kNN classifier and regressor;
Regression analysis: Linear regression, Ridge regression, Logistic regression; SVM, Naïve Bayes, Decision tree

Unsupervised Learning:
Introduction - Clustering - Evaluation metrics

Introduction to Neural Network: MP Neuron, Perceptron

TOTAL PERIODS: 45