

**Course Code:** PE-211

**Course Title:** Introduction to Data Sciences

## **Course Contents**

### **Basics of Data and its Analysis**

Basic data acquisition, cleaning, manipulation and pre-processing. Introduction to data analysis tools, Descriptive statistics, Data structures, Introduction to hypothesis testing and statistical inference.

### **Data Modeling and Visualization**

Data understanding and preparation, Linear Regression, Classification methods, including logistic regression, k-nearest neighbors, decision trees, and support vector machines, Exploratory data analysis and visualization. Implementing and validating linear and penalized regression, basic classification and basic clustering methods. Dimensionality reduction, including principle component analysis, Network analysis, Cleaning and reformatting messy datasets using regular expressions or dedicated tools such as open refine, Natural language processing, Big Data and its Ethics.

### **Text book**

1. Wes McKinney, “Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython”, 2nd Edition.

### **Reference Books**

1. Murphy, K., 2012, “Machine Learning: A Probabilistic Perspective”, MIT Press, Cambridge, MA.
2. Hands-On Machine Learning with Scikit-Learn and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems 1st Edition.