# Data Science and Machine learning with Python.

## **Course Duration: 3.5 months (14 weekends)**

# Section1: Course Introduction

* Introduction to the course
* What is Data Science?
* What is Machine learning?
* Environment setup and Installation.
* Jupyter Notebook Overview
* Spyder Overview
* Weka Tool for Data Analysis and ML

# Section 2. Python Cash Course

* Datatypes in Python (List / Dictionaries / Tuples)
* Functions, Procedure, lambda expressions, string slicing and dicing.
* Python comparison operators.
* Module Outline Loops and Conditionals statements.
* Datatypes in Python (List / Dictionaries / Tuples).
* Object Oriented Concepts in python. (Class / Object)
* Inheritance / Polymorphism.
* \*\*Python Crash course exercise {mini project}

Section 3. Statistics, Probability and Python Practice

* Types of Data
* Mean, Median, Mode
* Using mean, median, and mode in Python.
* Variation and Standard Deviation.
* Percentiles and Moments.
* Probability Density Function; Probability Mass Function.
* Outliers and its effects in model building.
* Inter Quantile Range \*(5-point summary)
* Covariance and Correlation
* Conditional Probability
* Bayes’ Theorem
* Exercise Solution: Conditional Probability of Purchase by Age

# Section4: Python Data Analysis

* Introduction to Numpy.
* Numpy array Indexing, Operations.
* Exercise on Numpy.
* Introduction to Pandas
* Pandas Series usage
* Dataframes in pandas and its usage
* Missing data treatment using pandas
* Groupby merging joining and concatenation operations using pandas.
* Introduction to Scipy library
* Matrix Operations using Scipy.
* Data Input and Output using Pandas.
* Connecting Database / XML and flat files via python.

**Project1**: Ecommerce Purchase order evaluations.

**Project** 2: SF salary

# Section 5: Python for Data Visualizations

* Introduction to Data Visualization section
* A Crash Course in matplotlib.
* Implementation of matplotlib on various datasets.
* Exercise on Matplotlib.
* A Crash course in Seaborn.
* Distribution plot
* Categorical plot
* Matrix plot
* Grids
* Regression plots
* Exercise on Seaborn’s lib.
* Pandas Built in Data Visualization.
* Plotly and Cufflinks
* Geographical plotting using Choropleth.

# Section 6: Introduction to Machine Learning.

* Introduction to machine learning.
* Supervised, Unsupervised and reinforcement learning.
* Classification vs clustering algorithms.
* **Linear Regression** model theory with mathematical Implementation.
* Linear regression with Python.
  + Exercise on Linear regression implementation using Scikit learn library.
* Project: Customer Analysis (Comparing the company website vs mobile application)
* Cross Validation and Bias variance Trade off.
* **Logistic Regression** model theory + mathematical implementation
* Where to use logistic regression, dataset analysis
* Logistic regression with Python.
  + Exercise on Logistic regression implementation using Scikit learn library.
  + Multivariate Regression, and Predicting Car Prices
* **K-Nearest Neighbors** + theory + Implementation with python
  + Mathematics behind K-Nearest Neighbors
  + Exercise on K-Nearest Neighbors implementation using Scikit learn library.
* **Decision Tree and Random forest** + theory + Implementation with python
  + Mathematics behind Decision tree
* What is **Ensemble method**?
  + Exercise on Decision Tree and Random forest classifier its implementation using Scikit learn library.
* **Support Vector Machines** + Maths behind Support vector machines.
  + SVM kernels
  + Linearly separable data
  + Non-Linearly separable data.
  + SVM project Overview
* **K-Means clustering** + hierarchical agglomerative clustering + mathematical implementation.
  + Exercise on K-means clustering
* **Dimensionality reduction** using Principal component analysis.
  + PCA with python
  + Exercise on PCA.

# Section7: Recommender system

* What is recommender system?
* Application of recommender system in real time applications.
* Types of recommender system (User based and Item based recommender system)
* Techniques to implement recommender system.
  + Exercise on recommender system with python.
  + **Project**: Movie recommendation for users.
    - Restaurant recommender system.

# Section 8: Natural Language processing. With Deep Learning

* Introduction to Natural Language processing
  + NLTK Python library.
  + Understanding Bag of words model for text analysis.
  + Learning the implementation of Lemmatization / stemming / other text processing techniques.
  + Email classification exercise.
  + Exercise on NLTK
* Neural Net and Deep Learning
  + What is tensorflow?
  + Tensorflow Installation.
  + Tensorflow basics.
  + MNIST with Multilayer perceptron example.
  + Tensorflow with Contrib Learn
  + TensorFlow Exercise.
* What is Keras?
  + Keras Basics.
  + Pipeline implementation using keras.
  + MNIST implementation with Keras.

**Project**

* Building your own Artificial Neural Network using Keras
* Building Convolution Neural Network (Image Recognition)

Section 9: Spark Overview

* Introduction to Spark.
* Learning Spark RDD’s.
* Accessing distributed application using pySpark

Section 10: Resume preparation