



Machine Learning with Python

(33 hours class room + 30 hours of practice sessions)

About the Course

Every day, around the United States, more than 36,000 weather forecasts are calculated. They gather all 36,000 forecasts, put them in a database, and compare them to the actual conditions encountered in that location on that day. All that collection, analysis, and reporting take a lot of heavy analytical horsepower and it is done with one programming language: Python. Over 40% of all data scientists use Python in their day to day work. Python has long been known as a simple programming language to pick up, which has propelled it to be the most preferred tool for a Data Scientist. In this course you will learn how to use the power of Python to analyze data, create beautiful visualizations, and use powerful machine learning algorithms to formulate business strategies.

Overview of the course

Class 1: Introduction to Python Programming Language

Introduction and Installation of Python software

Python packages: Pandas, & Numpy

Concepts of Data frame

Filtering

Loc and iloc for filtering

Usage of Boolean in Filtering

Appending

Class 2: Data handling in Python

Handling of Missing values

If else statement

Extra trick of using if else statement

Removal of Duplicates

Frequency Distribution

Merging – Inner, Outer, Left and Right

Binding and Appending

Descriptive Statistics

Inbuilt Numeric functions of R



Class 3: More data handling using Python

Pivot Table of Excel in Python

Grouping function

Learning of SQL queries using Python

Grouping numeric data

Class 4: Additional functions of Python

Text functions

Data cleaning with efficient text functions

Inbuilt String functions of Python

Reshape functions of Python

Class 5: Statistics

Everything you want to know about statistics....Well sort of!!

Mean, Median, Mode

Standard Deviation, Variance,

Normal Distribution

Hypothesis testing

T-test, Anova, Normality test

Class 6: Linear Regression

Predictive Analytics – Linear Regression

Concepts of Linear Regression

Simple and Multiple Linear Regression

Automatic Dummy Variables creation technique

Model Validation parameters

Model Assumption testing

Splitting of data for Validation and testing

Business Case Study with real data to model in Python



Class 7: Linear Regression Practice Case Study

Participants will be asked to develop a Linear Regression model on a real life data, in presence of the instructor. Time given is 2.5 hours. Participants will be treated like an industry employee, but in terms of help certainly the instructor will not be as ruthless as the boss. After completion of the model (with the help of the instructor wherever it is required), the instructor will show how to present a model to a real life client.

Class 7: Logistic Regression

Predictive Analytics – Logistic Regression
Concepts of Logistic Regression
Difference between Linear Regression and Logistic Regression
Automatic Dummy Variables creation technique
Model Validation parameters
Model Assumption testing
Splitting of data for Validation and testing
Business Case Study with real data to model in Python

Class 8: Logistic Regression Practice Case Study

Participants will be asked to develop a Logistic Regression model on a real life data, in presence of the instructor. Time given is 2.5 hours. Participants will be treated like an industry employee, but in terms of help certainly the instructor will not be as ruthless as the boss. After completion of the model (with the help of the instructor wherever it is required), the instructor will show how to present a model to a real life client.

Class 9: Time Series Forecasting

Time series forecasting: ARIMA
Difference between forecasting and prediction
Concepts of time series data
Concepts of ARIMA
Descriptive analytics for ARIMA
Development of model
Best model selection
Forecasting with the best model
Residual analysis
Business Case Study with real data to model in R software
Participants will be asked to develop a model in presence of the instructor.



Class 10: Cluster Analysis

Unsupervised Machine Learning with Python

Cluster Analysis: Concepts

Cluster analysis with Python – K Means, Hierarchical etc.

Class 11: Decision Tree and Random Forest

Concepts of Decision Tree

Decision Tree with Python

Concepts of Random Forest

Random Forest with Python

Important points:

1. After each class, assignments will be given as homework which are needed to be completed before the next class. The first 15 minutes of every class will be reserved to answer the participant's queries.
2. After every session, the discussed codes, presentations, handouts will be emailed to all the participants. Participants are advised to carry it either in soft copy or as print outs in the class.
3. Participants are advised to bring their own computers so that they can practice the codes along with the instructor.
4. Normally the class duration would be 3 hours, with a break of maximum 5-10 minutes depending of the requirement of the participants. In case all the queries of the participants are not answered with in the stipulated time of 3 hours then the instructor will extend the class by 15 minutes to 30 minutes.
5. After the completion of the module, there will be an option for all the participants to work on other case studies on real life data for further practice. (This is optional and will not be considered for calculating your final grade)
6. If a participant feels that he/she requires further help on certain topic, then they can attend the same session of some other batch.