Concepts Covered (Machine Learning with Python):

**Week 1:**

**Python for Data Science**:

Environment Setup and Basic Functionalities of Jupyter Notebook

Data type and Data Structure

Conditional Statements

Functions

**Data Manipulation**:

Series and Dataframe

Select and Filter

GroupBy, Aggregate and Sorting

Join and Concat

Handling missing values and Summarize functions

Data loading

**Data Visualization**:

Scatter plot

Histograms and density plot

Bar plot, Pie chart and box plot

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Project 1 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Week 2:**

**Linear Regression:**

Correlation, Covariance and R Square

Assumptions of Linear Regression

Residual Analysis

t-test and Anova

Multicollinearity

Heteroscedasticity

Autocorrelation

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Logistics Regression \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Week 3:**

**Logistics Regression:**

What is Logistics Regression

Assumptions and Hypothesis

Cost Function

Optimizing Cost Function

Performance Measure

Sensitivity, specificity and ROC curve

**Week 4:**

**Decision Tree:**

Decision Tree Modelling Objective

Basic Terminology

Algorithm behind Decision Tree

Entropy and Information Gain

Gini Index and Chi-Square

Bias and Variance

Tree Pruning

**Week 5:**

**Bagging and Boosting:**

Bagging

Majority Vote and OOB error

N\_estimators, Max\_features , Max\_depth

Feature Importance

Decision Tree Vs Random Forest

**Boosting:**

Weak learners

Cross-validation

XGBoost

Learning rate

Hyperparameter tuning – GridSearchCV

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Project 3 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Week 6:**

**Clustering:**

Distance Matrices

Hierarchical Clustering

K-means Clustering

Silhouette Score

Elbow Plot

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Project 4 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*