

## **Data Science Syllabus**

### **Basics to learn before**

40 – 100 hours

#### **Getting Started**

Local Setup and Development Environment

#### **Python Programming & Computer Science**

Types, Flow Control, Data Structures, Functions, OOP and Time Complexity

#### **SciPy Stack**

NumPy, pandas and matplotlib

#### **Mathematics**

Statistics, Probability, Calculus and Linear Algebra

### **Data Analysis**

100 - 160 hours

Exploratory data analysis - Getting, cleaning, analyzing and visualizing raw data is the main job responsibility of industry data scientists. Here you will learn how to discover patterns and trends that influence your future modeling decisions.

#### **Getting and Cleaning Data**

Static Files, SQL, Web Scraping, APIs and Messy Data

#### **Statistical Inference**

Event Space, Probability, Distributions and Hypothesis Testing

#### **Summarizing and Visualizing**

Data Descriptive Statistics, Univariate and Multivariate Exploratory Data Analysis

### **Machine Learning**

200 - 260 hours

Comprehensive set of machine learning algorithms from scratch, and master all the components of a predictive model, such as data preprocessing, feature engineering, model selection, performance metrics and hyperparameter optimization.

#### **Predictive Modeling**

Regression, Classification, Data Preprocessing, Model Evaluation and Ensembles

#### **Data Mining**

Dimensionality Reduction, Clustering, Association Rules, Anomaly Detection, Network Analysis and Recommender Systems

#### **Specialty Topics**

Data Engineering, Natural Language Processing, and Web Applications

### **Model Evaluation**

Different Model Evaluation Criteria

AUC ROC

RMSE

Logloss

### **Detailed Topics for Machine Learning**

#### **Learning Methodologies**

Supervised Learning

- \* Classification

- \* Regression

Unsupervised Learning

- \* Clustering

- \* Pattern Mining

Reinforcement Learning

Association Rules

Time series Analysis

Text Analysis (NLP)

Image Processing

#### **Exploratory Data Analysis**

Cleaning Data - Missing Values, Outliers

Preparing Data for Modeling - Transformations, Derived Variables

Visualization Methods and Applications in Excel

#### **Introduction to Inferential Statistics**

Understanding Probability and Distributions

Sampling Theory and How to Choose Representative Samples

Hypothesis Testing Concepts and Frameworks

Single Sample Hypothesis Tests - Z and T

Two Sample Tests - Independent and Paired

Multiple Samples Tests - ANOVA, Chi Square

Non-Parametric Tests

Case Study - HR Analytics

Case Study - Sales and Marketing Effectiveness

### **Linear Regression Models**

OLS Algorithm and Implementation in R

Model Building and Iterations with Linear Models

Interpretation of Output and Evaluating Model Results

Generating Business Insights and Outcomes from Linear Models

### **Logistic Regression Models and the MLE Algorithm**

Understanding the Odds Ratio

Building Logistic Models in R

Evaluating Logistic Regression Output - Probabilities, Confusion Matrix, Concordance, Lift

Generating Business Insights and Outcomes from Linear Models

### **Time Series Concepts**

Simple Exponential Smoothing

Holt-Winter's Forecasting

ARIMA

### **Bias, Variance and the bias-variance trade-off**

What is Bias in Machine Learning Models?

What is Variance in Machine Learning Models?

The Bias Variance Trade-off

## **Understanding the Machine Learning Approach to Algorithms**

Regression Trees

Classification Trees

Decision Trees

Random Forest and Ensemble Methods

Bagging and Boosting Algorithms

K-means Clustering

K-NN classification

Support vector machine

Naïve Bayes

Hierarchical clustering

Partitional clustering

Web scrapping

Object Recognition

## **Dimensionality Reduction Techniques**

Principal Component Analysis (PCA)

t-SNE

## **Interpretability of Machine Learning Models**

Different ways to interpret Machine Learning Models

LIME

## **Understanding Visualization and Storytelling Principles (Tableau)**

Information Hierarchy

The appropriate use of Color

Building interactive dashboards with Tableau

Building interactive dashboards with Power Bi

Creating an effective Story with Data

Visualisation with Tableau