

# SUPERVISED AND UNSUPERVISED LEARNING USING PYTHON

---



## SUPERVISED AND UNSUPERVISED LEARNING USING PYTHON

GOOGLE'S SELF-DRIVING CARS AND ROBOTS GET A LOT OF PRESS, BUT THE COMPANY'S REAL FUTURE IS IN MACHINE LEARNING, THE TECHNOLOGY THAT ENABLES COMPUTERS TO GET SMARTER AND MORE PERSONAL.

– ERIC SCHMIDT (GOOGLE CHAIRMAN)

This course is intended to give a holistic understanding on statistical & machine learning and its application using Python platform. The workshop will cover

- An introduction to business analytics
- An introduction to Python platform for data analysis
- An introduction to supervised machine learning algorithms
- An introduction to unsupervised machine learning algorithms
- Understanding of various sampling strategies and its efficacy in learning process
- An introduction to ensemble methods for handling imbalanced data
- Gradient descent algorithm and its application in finding the optimal solution
- Hands-on using the Python code and the real life dataset
- Introduction to different packages which can be used in Python for making robust and complex machine learning models

# SUPERVISED AND UNSUPERVISED LEARNING USING PYTHON

---

## OBJECTIVE

We are living in an era where computing moved from mainframes to personal computers to cloud. And while it happened, we started generating humongous amount of data. However the multi-folds increase in computing power also brought in advancement in application of algorithms which can be used to get insights from huge amount of data being generated. In this course, you will learn to nuances of building supervised and unsupervised machine learning models on real life datasets. We'll introduce you to Python platform and some of the statistical and machine learning algorithms which will become handy in solving challenging problems.

At the end of the course you will develop a clear understanding of the need of machine learning algorithms and the context in which to apply these algorithms to solve complex problems from the field of business.

## WHO SHOULD ATTEND

Irrespective of type of industry (retail, e-commerce, manufacturing, real estate & construction, telecom, hospitality, banking, healthcare, IT, supply chain & logistic, etc.); data forms the crux of decision making. This course is designed hone up analytical skills and business acumen of mid-level and senior level corporate professionals trying to understand the nuances of data science and help them the machine learning techniques an efficient way to generate insights for customers which in turn optimizes the bottom line of organizations.

## HARDWARE AND SOFTWARE

1. Participants should bring their laptop (preferably Windows 7 or higher/ Mac OS installed).
2. Operating System (any of the following):
  - Mac OS X with [XQuartz](#)
  - Windows (Version XP or later) is required.
3. Minimum 8 GB RAM on the system is advisable. 16 GB is preferable.

## INSTALLATIONS:

- For Windows, go to <https://docs.anaconda.com/anaconda/install/windows.html>
- For MacOS, go to <https://docs.anaconda.com/anaconda/install/mac-os>
- For Linux, go to <https://docs.anaconda.com/anaconda/install/linux>

More about anaconda can be found at <https://docs.anaconda.com>.

Participants are expected to resolve any installation issues of the software prior to the commencement of the session.

## PRE-REQUISITE & COURSE DELIVERABLE

1. Participants should have basic programming skills. Participants are expected to spend time with the code set as a home assignment to leverage the classroom training hours to the fullest.
2. High speed internet connection will be provided at the training venue.
3. Deliverable: Python code and dataset. Soft copy of the content being covered (PDF file)

# SUPERVISED AND UNSUPERVISED LEARNING USING PYTHON

---

## COURSE OUTLINE

### Day 1: Understanding Anaconda Framework platform and other useful packages in Python

#### Session 1–Introduction to Business Analytics

- What is Business Analytics
- Why is it needed and how industries are adopting it
- Different components of analytics
- Applications of analytics in different domains
- Different types of machine learning algorithms–Supervised, Unsupervised and Reinforcement learning

#### Session 2 & 3–Introduction to Anaconda and Python

- Overview of Anaconda framework
- Python – Variables, objects, loops, conditions, function.
- Python Data structures – lists, tuples, dictionaries, sets
- Introduction to Pandas – Data ingestion, descriptive statistics, visualization, frequent data operations, merging data frames

#### Session 4 & 5–Lab 1: Logistic Regression

- Introduction to logistic regression
- Logistic regression diagnostic: Classification Matrix, Sensitivity, Specificity, ROC Curve
- Strategy to find the optimal cut-off

### Day 2: Understanding supervised learning algorithms and its usage

#### Session 1–Lab 1: Logistic Regression...cont.

- Case study using logistic regression techniques
- Hands-on using Python code

#### Session 2–Lab 2: Decision Trees

- Decision tree – Classification and regression trees (CART), Gini Index
- Case study using decision tree techniques
- Hands-on using Python code

#### Session 3– KNN (K-Nearest Neighbors) and K-means

- Introduction to KNN algorithm
- Implementing KNN algorithm for imputation
- Introduction to clustering–K Means algorithm

# SUPERVISED AND UNSUPERVISED LEARNING USING PYTHON

---

- Hands-on using Python code for KNN and K-Means algorithm

## **Session 4–Machine Learning: Sampling Strategy**

- What is Machine learning
- Different sampling strategies–Bootstrapping, Up-Sample, Down-Sample, Synthetic Sample, Cross-Validation Data

## **Session 5–Lab 3: Machine Learning (Ensemble Methods)**

- Introduction to Bagging–Random Forest
- Introduction to Boosting– Adaptive boosting
- Case study of an imbalanced data and application of sampling strategies & ensemble methods
- Hands-on using Python code on an imbalanced data.

# SUPERVISED AND UNSUPERVISED LEARNING USING PYTHON

---

## COURSE SCHEDULE

### Day 1: Understanding Anaconda Framework platform and other useful packages in Python

This day will be primarily cover introduction to business analytics, introduction to Anaconda platform and regression concepts implementation using Python.

Topic	Session	From	To
Introduction to Business Analytics	1	9 AM	10:15 AM
Introduction to Python platform	2	10:30 AM	11:15 AM
Introduction to Python platform...cont.	3	12:00 PM	1:15 PM
Lab 1: Logistic regression	4	2:15 PM	3:30 PM
Lab 1: Logistic regression...cont.	5	3:45 PM	5:00 PM

### Day 2: Understanding machine learning and its implementation using Python

Day is primarily devoted to concept building on supervised and unsupervised machine learning and hands-on using Python code for the same

Topic	Session	From	To
Lab 2: Decision Trees	1	9 AM	10:15 AM
Lab 2: Decision Trees...cont.	2	10:30 AM	11: 45 AM
Lab 3: KNN (K-Nearest Neighbors) and K-means	3	12:00 PM	1:15 PM
Machine learning–Sampling strategies	4	2:15 PM	3:30 PM
Lab 4: Machine learning–Ensemble methods	5	3:45 PM	5:00 PM

# SUPERVISED AND UNSUPERVISED LEARNING USING PYTHON

---

## ABOUT INSTRUCTOR



**Rahul Kumar** is an alumnus of NIT Jaipur and IIM Bangalore. He has more than 14 years of experience spanning across software development, business consulting, analytical modelling and leading process improvement initiatives. He started his career in Information Technology sector and worked in Satyam Computers, Nokia Siemens and Deloitte Consulting before venturing into his own business. He co-founded a start-up ARIMA Research in June 2014 and was involved in internal operations and consulting engagement for various clients till December 2015. AwesomeStats Consulting is one of his new ventures, which is primarily focused towards trainings and consulting in the field of data science.

On the technical front, he works as a consultant at Indian Institute of Management Bangalore and has executed several analytics projects for large corporates. His recent work in the field of analytics includes predicting NPS for a reputed medical equipment manufacturer; predicting design issues for a leading US auto manufacturer; markdown optimization for a fashion retail client; sales and warranty forecasting for a leading auto manufacturer. He has also acquired a copyright on a research project funded by govt. of India, Ministry of Electronics and IT, on fraud analytics & credit scoring model for urban co-operative banks (ROC: SW-11742/2018) . He has taken more than 400 sessions in R/Python, as a guest faculty, at IIM Bangalore and IIM Lucknow and equivalent number of sessions for working professionals in various corporates. Few of the prominent corporate clients, he has worked with include General Electric, Cisco, Deloitte Consulting, United Health Group, HSBC, Flipkart, Fidelity Investments, General Motors, TVS Motors etc. He has also taken several faculty development programs in engineering and management institutions wherein he engaged with senior professors as well as with graduates and post graduates student. He has presented papers in several national and international conferences. Few of the prominent ones are:

- Rahul K., Seth N., Dinesh Kumar U. (2018) Spotting Earnings Manipulation: Using Machine Learning for Financial Fraud Detection. In: Bramer M., Petridis M. (eds) Artificial Intelligence XXXV. SGAI 2018. Lecture Notes in Computer Science, vol 11311. Springer, Cham
- Invited talk on “Using Machine Learning Algorithms to Detect Earnings Manipulations” at 5<sup>th</sup> International Conference on Business Analytics and Intelligence, IIM Bangalore 11<sup>th</sup>-13<sup>th</sup> December 2017.
- Paper on “Predicting Net Promoter Score (NPS) to Improve Patient Experience at Manipal Hospitals” published at Harvard Business Publishing, September 2017.
- Paper on “Behavioral Modeling to Predict Renege” published at Harvard Business Review, January 2016.
- Paper Presentation at CMMI conference organized by CMMI Institute, 10-11 Dec 2014 at Shenzhen, China.
- Paper Publication and Presentation at 6th International ITSM Conference organized by QAI Global Services in Bangalore, August 2013.
- Paper Presentation at SEPG Europe conference organized by SEI | Carnegie Mellon University, 5-7 June 2012 at Madrid, Spain.

He has also undergone workshop on the usage of statistical models and techniques from ISI Bangalore. His other certifications include DB2 certification from IBM and ISO 9001:2008 lead auditor certification by DNV India.