

Data Science for Everyone using Python (2 Days)

Data Science is emerging as a hot new profession and academic discipline. Harvard Business Review says Data Scientist is the Sexiest Job of the 21st Century. But demand for data scientists is racing ahead of supply. People with the necessary skills are scarce, primarily because the discipline is so new. This course is designed to give a start and introduction to this new discipline. This course is spread across 2 days and will have a plenty of hands on exercises on real world data.

Pre-requisites:

- ✓ A basic understanding of data and programming is required.
- ✓ Programming knowledge using Python is essential

Hardware & Software:

- ✓ A desktop or notebook with 64 bit OS (Windows/Mac)
- ✓ 8 GB RAM
- ✓ High speed Internet connection 256 kbps+

Topics:

- Setting up Data Analysis Environment in python
- Working with Numbers using Numpy
- Accessing, preparing and exploring data with Pandas & Scipy
- Data Exploration and visualization with Matplotlib
- Machine Learning using scikit-learn
- Making Predictions using Regression and classification algorithms
- Clustering
- Text Analytics

Days: 2 Days

Instructor Profile

Manaranjan Pradhan has about 15+ years of industry experience working on enterprise java, Cloud computing, Big Data and Data Mining. He has worked with TCS, HP, and iGATE patni and worked on large scale projects for customers like Motorola, Home Depot, CKWB Bank, P&G in the roles of solution and technical architect. He is a freelance who provides consulting and training on Cloud Computing, Big data & Hadoop and Data Science. He has been teaching Hadoop and data science for 2 years and has trained more than 500 people in Hadoop from large MNCs like EMC, CISCO, HP, YODLEE, YAHOO, SAMSUNG, VeriSign, Success Factors etc. He is also a guest lecture on Big Data Mining at IIM Bangalore.

Manaranjan Pradhan is a Cloudera Certified Developer for Apache Hadoop CDH4. He is an alumni of **Indian Institute of Management (IIM)**, Bangalore and has completed **certification on Business Analytics and Intelligence program**. He has **data science and scalable machine learning certifications from Coursera and edx.org**.

He writes his blog at <http://blog.enablecloud.com/>

Connect with him on Linked in <http://in.linkedin.com/pub/manaranjan-pradhan/a/6bb/314>

Data Science for Everyone using Python (2 Days)

Introduction to Data Science and Setting up data analysis environment	Introduction to Data Science Setting up Python Environment for Data Analysis Overview of Data Analysis Software Stack - Numpy, Pandas, Matplotlib, Scipy and Scikit-learn Hands On Exercise
Working with Numbers	Introduction to Numpy array Overview of Array and operations N-Dimensional array and manipulations
Accessing and preparing data with Pandas	Loading data from varieties of sources: CSV, Databases Data manipulation - Filtering, Grouping, Ordering of data Dealing with missing Data Dealing with Continuous and categorical variables Normalizing and transforming data
Data Exploration and Visualization using Scipy and matplotlib	Basic Statistical analysis using Scipy Univariate Analysis – Statistics Summary, Hypothesis tests Drawing Histograms, Bar charts, Density Plots, Box Plots Drawing Density plots and understating data distributions
Regression and Classification Algorithms	Understand Regression Techniques Simple Linear Regression & Multiple Linear Regressions Measuring accuracy of the models Regression Diagnostics - Validating Models Making Predictions using Classification algorithms - Logistic Regression
Clustering	Understanding k-means clustering and creating Segments Creating clustering plots and Dendograms Creating customer segments using Clustering Algorithms
Text Analytics	Handling Text and unstructured Data Accessing Social Data - Integrate with Twitter Trend Analysis & Topic Modelling
Emerging areas and next steps	Emerging Opportunities and Challenges of Big Data Roles & Responsibilities of Data Scientists