

DATA SCIENCE

MODULE 1 : GETTING STARTED WITH PYTHON LIBRARIES

what is data analysis ?
why python for data analysis ?
Essential Python Libraries
Installation and setup
Ipython
Jupyter Notebook

MODULE 2 : NUMPY ARRAYS

Creating multidimensional array
NumPy-Data types
Array attributes
Indexing and Slicing
Creating array views and copies
Manipulating array shapes
I/O with NumPy

MODULE 3 : WORKING WITH PANDAS

Installing pandas
Pandas dataframes
Pandas Series
Data aggregation with Pandas DataFrames
Concatenating and appending DataFrames
Joining DataFrames
Handling missing data

MODULE 4 : DATA LOADING, STORAGE AND FILE FORMAT

Writing CSV files with numpy and pandas
HDF5 format
Reading and Writing to Excel with pandas
JSON data
Parsing HTML with Beautiful Soup
PyTables

MODULE 5 : STATISTICS AND LINEAR ALGEBRA

Basic statistics with numpy
Linear Algebra with numpy
Numpy random numbers
Creating a numpy masked array

MODULE 6 : DATA VISUALIZATION

Installation matplotlib
Basic matplotlib plots
Scatter plots
Saving plots to file
plotting functions in pandas

MODULE 7 : INTRODUCTION TO MACHINE LEARNING

What is ML..?
Types of ML
Decision trees
Linear regression

Logistic regression
Naive Bayes
k-Nearest Neighbors

MODULE 8 : NATURAL LANGUAGE PROCESSING

Install nltk
Tokenize words
Tokenizing sentences
Stop words with NLTK
Stemming words with NLTK
Speech tagging
Sentiment analysis with NLTK

MODULE 9 : INTRODUCTION TO OPENCV

Setting up opencv
Loading and displaying images
Applying image filters
Tracking faces
Face recognition

MODULE 10 : WORKING WITH BIG DATA

What is Hadoop?
MapReduce
File handling with Hadoopy
Pig
Pyspark

MACHINE LEARNING

MACHINE LEARNING **INTRODUCTION TO MACHINE** **LEARNING**

What is Machine learning?
Overview about sci-kit learn and tensorflow
Types of ML
Some complementing fields of ML
ML algorithms
Machine learning examples

MODULE 2: **REGRESSION BASED LEARNING**

Simple regression
Multiple regression
Logistic regression
Predicting house prices with regression

MODULE 3: **CLUSTERING BASED LEARNING**

Definition
Types of clustering
The k-means clustering algorithm

MODULE 4: **DATA MINING**

Introducing data mining
Decision Tree
Affinity Analysis
Clustering

MODULE 5: **CLASSIFICATION – SENTIMENT ANALYSIS**

MODULE 6 **NATURAL LANGUAGE PROCESSING**

Install nltk
Tokenize words
Tokenizing sentences
Stop words with NLTK

Stemming words with NLTK
Speech tagging
Sentiment analysis with NLTK

MODULE 7: **MAKING SENSE OF DATA THROUGH** **VISUALIZATION**

Introducing matplotlib
Bar Charts
Line Charts
Scatter plots
Bubble charts

MODULE 8: **WORKING WITH OPENCV**

Setting up opencv
Loading and displaying images
Applying image filters
Tracking faces
Face recognition

MODULE 9: **PERFORMING PREDICTIONS WITH LINEAR** **REGRESSION**

Simple linear regression
Multiple regression
Training and testing model

MODULE 10: **SUPPORT VECTOR MACHINES(SVM)**

MODULE 11: **NEURAL NETWORKS**