

DATA SCIENCE MASTER PROGRAM

Course Curriculum



Getting Trained in Data Science Course

is never hard with us.

www.i2tutorials.com

6 Modules : 150 Hours : 3 Months : 3 Capstone Projects

Just be a Data Science Professional, shape your Career now....

- **Statistics – Mathematics**
- **Machine Learning**
- **R & Python Programming**
- **Artificial Intelligence using Tensor Flow**
- **Machine Learning using AWS, Azure**
- **BigData Analytics**

Understanding the Data

- Data, Data Types
- Meaning of variables
- Central Tendency
- Measures of Dispersion
- Measures of Variability
- Measures of Shape
- Data Distribution
- Correlation, Covariance
- Practical Examples

Probability Distributions

- Mean, Expected value
- Binomial Random Variable
- Normal Distribution
- Poisson Random Variable
- Continuous Random Variable
- Discrete Random Variable
- Practical Examples

Sampling Distributions

- Central Limit Theorem
- Sampling Distributions for Sample Proportion, p -hat
- Sampling Distributions for Sample Mean, \bar{x}
- Z- Scores
- Practical Examples

Hypothesis Testing

- Type I and Type II Errors
- Decision Making
- Power
- Testing for mean, variance, proportion
- Practical Examples

Association between Categorical Variables

- Contingency Tables
- Independent and Dependent
- Pearson's Chi-Square Test
- Misuses of Chi-Squared Test
- Measures of Association
- Practical Examples

ANOVA Analysis

- Analysis of Variance & Co-Variance
- ANOVA Assumptions & Comparisons
- F-Tests
- Practical Examples

Matrices

- One – Dimensional
- Multi Dimensional
- Arithmetic Operations
- Examples

Vectors

- Understanding Vectors
- Scalar Vs Vector
- Arithmetic Operations
- Examples

Machine Learning

Module – 2 (40 hrs)

Supervised Learning

- An Approach to Prediction
- Least Squares and Nearest Neighbors
- Statistical Decision
- Regression Models

Linear Methods for Regression

- The Gauss–Markov Theorem
- Multiple Regression
- Forward- and Backward-Stepwise Selection
- Ridge Regression
- Lasso Regression
- Example using R / Python

Linear Methods for Classification

- Linear Regression of an Indicator Matrix
- Linear Discriminant Analysis
- Logistic Regression
- Rosenblatt’s Perceptron Learning Algorithm
- Example using R / Python

Kernel Smoothing Methods

- One-Dimensional Kernel Smoothers
- Local Linear Regression
- Local Polynomial Regression
- Mixture Models for Density Estimation and Classification
- Example using R / Python

Model Selection

- Bias, Variance and Model Complexity
- Optimism of the Training Error Rate
- Vapnik–Chervonenkis Dimension
- Cross-Validation

Model Inference & Averaging

- Bootstrap and Maximum Likelihood Methods
- Relationship Between the Bootstrap and Bayesian Inference
- The EM Algorithm
- Bagging
- Example using R / Python

Tree-Based Methods

- Regression Trees
- Classification Trees
- Bump Hunting
- MARS: Multivariate Adaptive Regression Splines
- Example using R / Python

Boosting

- Steepest Descent
- Gradient Boosting
- Regularization
- Interpretation
- Example using R / Python

Neural Networks

- Fitting Neural Networks
- Over fitting
- Hidden Units
- Multiple Minima
- Single, Multi-Layer Perceptron
- Example using R / Python

Support Vector Machines (SVM)

- Support Vector Classifier
- Generalizing Linear Discriminant Analysis
- Flexible Discriminant Analysis
- Penalized Discriminant Analysis
- Example using R / Python

K-Nearest-Neighbor Classifiers

- Prototype Methods
- K-means Clustering
- Vector Quantization
- Gaussian Mixtures
- k-nearest Neighbors
- Example using R / Python

Unsupervised Learning

- The Apriori Algorithm
- Unsupervised as Supervised Learning
- Generalized Association Rules
- K-means Cluster Analysis
- Hierarchical Clustering
- Principal Components, Curves and Surfaces
- Non-Linear Dimension Reduction
- The Google Page Rank Algorithm
- Example using R / Python

Random Forests

- Variable Importance
- Random Forests and Over fitting
- Bias
- Adaptive Nearest Neighbors
- Example using R / Python

Python

Module -3 (20 hrs)

Introduction to Python

- Installation of Python framework and packages: Anaconda & pip
- Working with Jupyter notebooks
- Creating Python variables
- Numeric , strings
- logical operations
- Lists
- Dictionaries
- Tuples
- sets
- Practice assignment

Iterative Operations & Functions in Python

- Writing for loops in Python
- While loops and conditional blocks
- List/Dictionary comprehensions with loops
- Writing your own functions in Python
- Writing your own classes and functions
- Practice assignment

Data Handling in Python using Packages

- Numpy
- Pandas
- SymPy
- SciPy
- Matplotlib

Data Visualization in Python

- Need for data summary & visualization
- Summarising numeric data in pandas
- Summarising categorical data
- Group wise summary of mixed data
- Basics of visualisation with ggplot & Seaborn
- Inferential visualisation with Seaborn
- Visual summary of different data combinations
- Practice assignment

Data preparation using Python

- Needs & methods of data preparation
- Handling missing values
- Outlier treatment
- Transforming variables
- Data processing
- Practice

R - Programming

Module -4 (20 hrs)

Fundamentals of R

- Installation of R & R Studio
- Getting started with R
- Basic and Advanced Data types in R
- Variable operators in R
- Working with R data frames
- Reading and writing data files to R
- R functions and loops
- Special utility functions
- Merging and sorting data
- Practice assignment

Univariate statistics in R

- Summarizing data, measures of central tendency
- Measures of data variability & distributions
- Using R language to summarize data
- Practice assignment

Data visualization in R

- Introduction exploratory data analysis
- Descriptive statistics, Frequency Tables and summarization
- Univariate Analysis (Distribution of data & Graphical Analysis)
- Bivariate Analysis (Cross Tabs, Distributions & Relationships, Graphical Analysis)
- Creating Graphs (Bar/pie/line chart/histogram/boxplot/scatter/density etc)
- R Packages for Exploratory Data Analysis (dplyr, plyr, gmodels, car, vcd, Hmisc, psych, doby etc)
- R Packages for Graphical Analysis (base, ggplot, lattice,etc)

Hypothesis testing and ANOVA in R

- Introducing statistical inference
- Estimators and confidence intervals
- Central Limit theorem
- Parametric and non-parametric statistical tests
- Analysis of variance (ANOVA)

Data preparation using R

- Needs & methods of data preparation
- Handling missing values
- Outlier treatment
- Transforming variables
- Data processing with dplyr package
- Practice

Artificial Intelligence

Module – 5 (25 hrs)

Introduction to Deep Learning

Deep Learning: A revolution in Artificial Intelligence
Limitations of Machine Learning
Deep Learning vs Machine learning
Examples of Deep Learning
Implementations where Deep Learning is applicable

Glance of Machine Learning Algorithms

Regression
Classification
Clustering
Reinforcement Learning
Underfitting and Overfitting
Optimization

Understanding Fundamentals of Neural Networks with Tensorflow

How Deep Learning Works?
Activation Functions
Illustrate Perceptron
Training a Perceptron
Important Parameters of Perceptron

Tensor Flow

What is TensorFlow?
Use of TensorFlow in Deep Learning
Working of TensorFlow
How to install Tensorflow
HelloWorld with TensorFlow
Tensorflow code-basics
Graph Visualization
Constants, Placeholders, Variables
Creating a Model
Running a Machine learning algorithms on TensorFlow

Deep dive into Neural Networks with Tensorflow

Understand limitations of A Single Perceptron
Neural Networks in Detail
Multi-Layer Perceptron
Backpropagation – Learning Algorithm
Understand Backpropagation – Using Neural Network Example
MLP Digit-Classifer using TensorFlow
TensorBoard

Convolutional Neural Networks (CNN)

Define CNNs
Discuss the Applications of CNN
Explain the Architecture of a CNN
List Convolution and Pooling Layers in CNN
Illustrate CNN

Transfer Learning of CNNs

Introduction to CNNs
CNNs Application
Architecture of a CNN
Convolution and Pooling layers in a CNN
Understanding and Visualizing a CNN
Transfer Learning and Fine-tuning Convolutional Neural Networks

Recurrent Neural Networks (RNN)

Intro to RNN Model
Applications of RNN
Modelling sequences
Training RNNs with Backpropagation
Long Short-Term memory (LSTM)
Recursive Neural Tensor Network Theory
Recurrent Neural Network Model

Hands-On Project

Machine Learning in Cloud & Big Data Analytics Module -6 (20 hrs)

Machine Learning using Azure

- What is Microsoft Azure?
- Azure Machine Learning
- Diving into Azure Machine Learning
- Training a Model
- Deploy a Model
- Practical Example

Machine Learning using AWS

- What is AWS?
- AWS Machine Learning
- Diving into AWS Machine Learning
- Training A Model
- Deploy a Model
- Practical Example

Working with Mode Analytics

- What is Mode Analytics?
- Data Science on Web Model
- Using SQL in mode analytics
- R note book in mode analytics
- Python note book in Mode analytics
- Working on Practical Example

Introduction to Big Data analytics

- Hadoop – HDFS
- Mapreduce
- Hive
- Hbase
- Spark
- Spark SQL
- Spark Mlib

Machine Learning using Spark

- Introduction
- Data sets
- Data frames
- Machine Learning using spark
- SparkR
- PySpark
- Practical Example