BECOME AN EXPERT IN

DATA SCIENCE

StatiSticS | ai | ML | DL | NLP | R | PythoN| Tensor fl ow

About the Course

Data Science is the study of the generalizable extraction of knowledge from data. Being a data Scientist requires an integrated skill set spanning mathematics, statistics, machine learning, databases and programming languages along with a good understanding of the craft of problem formulation to engineer effective solutions.

This course will introduce students to this rapidly growing field and equip them with some of its basic principles and tools as well as its general mindset.

- Students will learn concepts, techniques and tools they need to deal with various facets of data science practice, including data collection and integration, exploratory data analysis, predictive modeling, descriptive modeling, data product creation, evaluation, and effective communication.
- The focus in the treatment of these topics will be a balanced approach on breadth and depth, and emphasis will be placed on integration and synthesis of concepts and their application to real time problems.
- To make the learning contextual, real datasets from a variety of disciplines will be used.

Program Highlights

- ✓ Most Comprehensive Curriculum
- ✓ Trained by passionate and Industry experts
- ✓ Each concept will be explained by golden rule

  Theory → Example → Software Implementation (R/Python)→ Real-Time applicability
- ✓ All classes explained with REAL TIME projects experience
- ✓ End to End Data Science Project work flow explanation
- ✓ Designed for the Industry
- ✓ Live Project
- ✓ Placement Assistance
- ✓ Free Mock Interviews for Data Science Interview preparation
- ✓ Hand written notes copy and slides copy
- ✓ Detailed assistance in Resume preparation. Special attention for experienced people on previous experience
- ✓ Latest resources, blogs and articles sharing

Audience

Any degree. No programming and Statistics knowledge required.

Duration & Mode of Training

- ✓ 3 months, Online Training

Course Content

Introduction to Data Science

- What is data science?
    - ✓ How is data science different from BI and Reporting?
- What is difference between AI, Data Science, Machine Learning, Deep Learning
- Job Land scape and Preparation Time
- Who are data scientists?
    - ✓ What skillsets are required?
- What is day to day job of Data Scientist
    - ✓ What kind of projects they work on?
- End to End Data Science Project Life Cycle
- Data Science roles – functions, pay across domains, experience

Business Statistics

- Introduction to statistics
- Summarizing Data
    - Central Tendency measures – Mean, Median and Mode
    - Measures of Variability – Range, Interquartile Range, Standard Deviation and Variance
    - Measures of Shape – Skewness and Kurtosis
    - Covariance, Correlation
- Data Visualization
    - Histograms
    - Pie charts
    - Bar Graphs
    - Box Plot
    - Scatter plot
- Probability basics
- Parametric and Non parametric Statistical Tests

- 'f' Test
- 'z' Test
- 't' Test
- Chi-Square test
- Probability Distributions
  - Expected value and variance
  - Discrete and Continuous
  - Bernoulli Distribution
  - Binomial Distribution
  - Normal Distribution
  - Uniform Distribution
  - Empirical Rule
  - Chebyshev's Theorem
- Sampling methods and Central Limit Theorem
  - Overview
  - Random sampling
  - Stratified sampling
  - Cluster sampling
  - Central Limit Theorem
- Hypothesis Testing
  - Type I error
  - Type II error
  - Null and Alternate Hypothesis
  - Reject or Acceptance criterion
  - P-value
- Confidence Intervals
- ANOVA
  - Assumptions
  - One way
  - Two way

Artificial Intelligence – Machine Learning Introduction

- Introduction to Machine Learning
  - What is Machine Learning?
  - Statistics (vs) Machine Learning
  - Types of Machine Learning
    - Supervised Learning
    - Un-Supervised Learning

- Reinforcement Learning

Artificial Intelligence – Supervised Machine Learning

- Classification
    - Nearest Neighbor Methods (knn)
    - Logistic
- Tree based Models – Decision Tree
    - Basics
    - Classification Trees
- Probabilistic methods
    - Bayes Rule
    - Naïve Bayes
- Regression Analysis
    - Simple Linear Regression
    - Assumptions
    - Model development and interpretation
    - Sum of Least Squares
    - Model validation
    - Multiple Linear Regression
- Regression Shrinkage Methods
    - Lasso
    - Ridge
- Advanced Models – Black Box
    - Support Vector Machine
    - Neural Networks
- Ensemble Models
    - Random Forest
    - Gradient Boosting
- Optimization
    - Gradient Descent (Batch and Stochastic)

Artificial Intelligence – Unsupervised Machine Learning

- Association Rules (Market Basket Analysis)
    - Apriori
- Cluster Analysis
    - Hierarchical clustering
    - K-Means clustering

- Dimensionality Reduction
    - Principal Component Analysis

Model Validation

- Confusion Matrix and its metrics
- ROC Curve (AUC)
- R Squared
- Adjusted R Squared
- Root Mean Square Error (RMSE)
- K-fold Cross Validation

Artificial Intelligence – Natural Language Processing

- What is Natural Language Processing
- Applications of NLP
- Setting Environment
- Text Cleaning, Pre-processing techniques
    - ✓ Tokenization
    - ✓ Stemming
    - ✓ Lemmatization
    - ✓ Stop words
    - ✓ POS Tagging
    - ✓ Named Entity Recognition
    - ✓ Count Vectorizer
    - ✓ Term Frequency (TF) - Inverse Document Frequency (IDF)
- Text Classification - Project #1
- Text Classification - Project #2
- Information Retrieval System - Project #2

Artificial Intelligence – Deep Learning Introduction

- Deep Learning: A revolution in Artificial Intelligence
- Limitations of Machine Learning
- What is Deep Learning?
- Advantage of Deep Learning over Machine learning
- 3 Reasons to go for Deep Learning
- Real-Life use cases of Deep Learning

Recurrent Neural Network

- Introduction to RNN Model
- Why RNN as against Feed Forward Networks
- Application use cases of RNN
- Modeling sequences
- Training RNNs with Back propagation
- Recurrent Neural Network Model

Convolutional Neural Network

- Introduction to CNNs
- CNNs Application
- Architecture of a CNN
- Convolution and Pooling layers in a CNN
- Understanding and Visualizing a CNN

R Programming Language

- Introduction
    - R Overview
    - Installation of R and RStudio software
    - Important R Packages
    - Datatypes in R – Vectors, Lists, Matrices, Arrays, Data Frames
- Decision making & Loops
    - If-else, while, for
    - Next, break. try-catch
- Functions
    - Writing functions
    - Nested functions
- Built-in functions
    - Apply, Sapply , Lapply etc.
- Data Preparation/Manipulation
    - Reading and Writing Data
    - Summarize and structure of data
    - Exploring different datasets in R
    - Sub Setting Data Frames
    - String manipulation in Data Frames

- Handling Missing Values, Changing Data types, Data Binning Techniques, Dummy Variables
- Data Visualization
  - Basic charts – Histograms, Bar plots, Line graphs, Scatter plots etc.

Python Programming Language

- Introduction
  - How is Python different from R
  - Installing Anaconda- Python
  - Setting up with spyder
- Datatypes in Python
- Importing modules
- Control loops
- Numpy
- Pandas
- Scikit-Learn - Machine Learning in Python
- Matplotlib

Tensor Flow

- What is Tensor Flow?
- History & Evolution
- Architecture
- Tensor Flow Installation & Set-up
- Tensor Flow code-basics
- Graph Visualization
- Constants, Placeholders, Variables
- Creating a Model
- Step by Step - Use-Case Implementation for the following
  - Supervised Machine Learning - Classification & Regression
  - Natural Language Processing - Text Classification
  - Computer Vision - Image Classification