



# Data Science & Artificial Intelligence

## *Course contents*

### Module 1: Introduction to Data Science

- What is Data Science?
- Skillsets required for Data Scientists
- Data Science Process
- Standard Lifecycle of Data Science Projects
- Job opportunities and demand for Data Scientists
- What is Business Intelligence
- What is Data Mining
- What is Analytics
- ~~Types of Analytics~~ -----
- Data Science Roles, Responsibilities , Jobs and Market Demand
- What is Machine Learning
- What is Deep Learning
- What is AI

### Data

- What is Data
- Types of Data
- Data collection types

- Data Architecture
- Components of Data Architecture

## **Module 2: PYTHON for Data Science**

### **❖ Python programming for Data Science**

- Python Environment Setup and Essentials
- Anaconda & Jupyter Notebook Installation
- Variable Assignment, operators, Data types
  
- Indexing & Slicing
- Data structures: Lists, Tuples, Sets, Dictionaries
- Functions
- Conditional flow statements: If, For, While
- Map, Filter and Reduce functions
- Lambdas and List Comprehensions

### **❖ Numerical Computing using NumPy**

- ndarray: Purpose, Properties, Types, Axis
  - creating 1d, 2d and 3d arrays
  - Accessing Array Elements
  - Indexing, Slicing, Iteration, with Boolean and Integer Arrays
  - Array manipulation
  - Linear Algebra using Numpy
-

## ❖ Data Analysis using PANDAS

- Understanding Pandas
- Defining Data Structures: Series, Dataframes, Panels
- Working with Series and Data Frames
- DataFrame operations
- Indexing: .loc and .iloc
- DataFrame functions: pipe/apply/applymap
- Data Analysis: case study1: Importing and exporting data
- Case study2: Cleaning data [filtering, removing duplicates etc]
- Case study3: Handling missing values
- Case study4: Data wrangling
- Case study5: Grouping and Aggregation
- Case study6: merging, joining, concatenation

## ❖ Data Visualization using Matplotlib & Seaborn

- Features of Matplotlib
- Working with various Plots

**Capstone Project 1: EA Games: Customer & Game review analysis**

**Capstone Project 2: MovieLens data analysis**

## Module 3: STATISTICS

### ❖ Descriptive Statistics

- Variables in Statistics
- Measuring the Central Tendency – Mean, Median, mode, Range, Quartiles
- Measuring Spread – Variance and Standard Deviation
- Understanding Numeric Data – Uniform and Normal Distributions
- Probability Refresher
- Probability density functions

Central Limit Theorem



## ❖ Hypothesis Testing & Inferential Statistics

- Importance of Hypothesis Testing in Business
- Null and Alternate Hypothesis
- Type 1 and Type 2 Errors
- Significance level and Power
- Upper Tail Test and Test Statistics
- Z-Test, t-Test and F test
- Chi-Square Test
- ANOVA
- Correlation and covariance
- Linear Regression, Logistic regression

## Module 4: R for Data Science

### R essentials

- Introduction to R
- Installing R and R packages
- R studio
- R Programming
- Operators in R: Arithmetic,  
Relational,  
Logical,  
Assignment
- R Data Structures – Vectors,  
Factors,  
Lists,  
Data  
Frames,

## Arrays

- Conditional Statements and loops in R
- R Functions: Commonly Used and String Functions
- Working on Files in R: Importing data from Files?  
Importing Table and CSV Files
- Understanding Dplyr Package
- Managing Data with R
- Statistical Modelling and Data Mining using R
- Understanding Types of Data: Qualitative Analysis,  
Quantitative Analysis,  
Univariate data  
analysis Bi variate  
data analysis  
-----  
Multi variate data analysis -----

## **Module 5: Exploratory Data Analysis [EDA]**

- What is EDA
- Goals of EDA
- Introduction to Statistical Plots
- Visualizing Numeric Variables
- Visualizing Categorical variables
- One Dimensional Charts
- Histograms
- Bar Charts
- Two Dimensional Charts
- Visualizing Relationships – Scatterplots
- Box Plots
- Multi-Dimensional Plots

### **Capstone Project 3: Iris EDA**

### **Capstone Project 4: Habermans Cancer survival analysis**

  
-----

## **Module 6: MACHINE LEARNING**

- ❖ **Introduction to Machine Learning using Scikit Learn**
  - What is Machine Learning?

- How do Machines Learn?
- Abstraction and Knowledge Representation
- Generalization
- Steps to apply Machine Learning to your Data
- Choosing a Machine Learning Algorithm
- Introduction to Types of Machine Learning Algorithms

## ❖ **Supervised Learning Techniques and Algorithms**

- Steps in Supervised Learning Techniques and Algorithms
- Understanding Process Flow of Supervised Learning Techniques
- Training, Validation and Testing
- Regression
- Gradient Descent
- Classification
- Measures of Performance
- R-Square and RMSE
- Confusion Matrix
- Accuracy, Precision and Recall
- F-Score
- ROC curve (Receiver Operating Characteristic curve)
- Bias – Variance tradeoff
- Underfitting and Overfitting
- Understanding Classification and Prediction
- K-NN, Naïve Bayes, Support Vector Machines
- Decision Trees and Random Forests

## ❖ **Unsupervised Learning Techniques & Algorithms**

- Studying Clustering
- Understanding K-means Clustering
- What is Hierarchical Clustering?
- Hierarchical Clustering Algorithm
- Association Rule Mining
- Checking Apriori Algorithms

**Capstone Project 5: House Prices**

**Prediction Capstone Project 6: Titanic**

**Survival analysis**

**Capstone Project 7: Diabetes prediction [Healthcare Analytics]**

**Capstone Project 8: Predicting Gas consumptions**

**Capstone Project 9: Iris classification**

**Capstone Project 10: Customer Segmentation [Clustering]**

---

### **Module 7: Deep Learning and Computer Vision**

- Understanding Neural Networks
- Network Topology
- Neural Networks: Master Feed-Forward
- Recurrent and Gaussian Neural Network
- Training Neural Networks with Backpropagation
- Artificial Neural network
- Recurrent Neural Network
- Introduction to Computer Vision
- Convolution neural network
- Transfer Learning
- Introduction to Tensorflow and Keras
- Building Neural network using Tensorflow

**Capstone Project 11: Object detection**

---

### **Module 8: Natural Language Processing (NLP)**

- NLP Environment Setup & Applications
- NLP Sentence Analysis & Libraries
- NLTK
- Lemmatization
- Stemming
- Topic modelling
- Entity Recognition

**Capstone Project 12: Text classification**

---

**Module 9: Tableau [with project]**

**Module 10: Introduction to Hadoop & Spark**

**Bonus (additional)**

**Projects QL Loan**

**prediction case study**

**Investment Banking case**

**study Uber supply demand**

**gap analysis Stock market**

**prediction**

**Bike rentals prediction**