DATA SCIENCE

Module - 1 (Python Basics)

Welcome To The Course

- What is Artificial Intelligence
- Introduction To DataScience
- Real Time UseCases Of DataScience
- Who is a DataScientist??
- DataScience Project Lifecycle
- Skillsets needed for DataScientist
- Difference between DataEngineer, DataScientist and DataAnalyst
- 6 Steps to take in 3.5 Months for a complete transformation to DataScience from any other domain
- Machine Learning-Giving Computers The ability to learn from data
- Supervised vs Unsupervised
- DeepLearning vs Machine Learning

Python Fundamentals

- Software Installation
- Jupyter Notebook Tutorial
- Introduction to Python
- Comments
- Variable,Operators,DataTypes
- If Else,For and While Loops
- Functions
- Lambda Expression
- Taking input from keyboard
- List
- Tuple
- Set
- Dictionary
- INTERVIEW QUESTIONS ASSIGNMENT-1

Module - 2 (Python Advance)

NumPy

- Introduction to Numpy

- Creating Arrays
- arange(),linspace() etc.
- Creating Arrays of Random Numbers
- Basic Operations on an Array
- Applying Universal functions on an array
- Linear Algebra operations on an array
- Numpy DataTypes
- Type Conversion
- Array Stacking
- ASSIGNMENT-2

Scipy

SciPy Intro

SciPy Constants

SciPy Optimizers

SciPy Sparse Data

SciPy Graphs

SciPy Spatial Data

Pandas

- Introduction to Pandas
- Creating DataFrames
- Reading data from csv,excel etc. into a DataFrame & writing df into csv,excel
- Selection and Indexing
- Conditional Selection
- Groupby
- Pivot Table
- Merging , Joining, Cancatenation
- Missing Value Treatment
- Data Visualisation using Pandas
- ASSIGNMENT-3

Module - 3 (Visualisation)

Visualisation-Plotly

- Line Plots
- Scatter Plots
- Pair Plots
- Histograms
- Heat Maps
- Bar Plots
- Count Plots
- Factor Plots
- Box Plots
- Violin Plots
- Swarm Plots
- Strip Plots
- Pandas Builtin Visualisation Library
- ASSIGNMENT-4

Module - 4 (Statistics)

Statistics

- Descriptive vs Inferential Statistics
- Mean,Median,Mode
- Central Limit Theorm
- Measure of dispersion
- Inter Quartile Range
- Variance
- Standard Deviation
- Box Plot
- Z score
- Scatter Plot
- Pearson’s Product Moment Correlation-r
- R square
- Adjusted R-square
- Normal Distribution
- Standard Normal Distribution
- Emprical rule of Normal Distribution
- What is an Outlier
- Outlier Detection and Removal

MACHINE LEARNING

Module - 5 (ML-Linear Reg)

Linear Regression, Cost Function, Gradient Descent, Sklearn

- Introduction to Machine Learning
- Supervised vs Unsupervised
- Regression vs Classification
- Linear Regression Theory
- Gradients/Derivative Theory
- Assumption of Linear Regression
- Cost Function
- Optimize Cost function using Gradient Descent
- Gradient Descent Detail Explanation
- Mathematical Derivation
- Multi- Colinearity
- MAE
- MSE
- RMSE

Module - 6 (Decision Tree, Random Forest)

Decision Tree

- What is ID3 Algorithm
- Entropy
- Calculating Information Gain
- Overfitting, Underfitting, Best fit

Random Forest

- What is Bootstap
- Bagging
- Difference between Random Forest and Decision Tree
- Feature Selection using Random Forest
- Hyperparameter tuning

CLASSIFICATION VALIDATION TECHNIQUES

- Confusion Matrix
- Classification Report
- Recall
- Precision

- AUC
- ROC
- Cross Validation

Module - 7 (PCA)

Principal Component Analysis

- Introduction to Dimensionality Reduction
- PCA Theory discussion
- Eigen Values , Eigen Vectors
- Step by Step Detail Mathematical Derivation
- Individual and Cummulative Variation Ratio

Step By Step Implementation of PCA From scratch (with out sklearn) and by using sklearn

- Implement PCA from scratch using Numpy and using sklearn is a real time dataset

Module - 8 (KMeans)

KMeans Clustering

- Introduction to Unsupervised Machine Learning
- KMeans Theory
- How to decide K in KMeans

Module - 9 (NLP)

Text Mining

- Introduction to NLP
- Text Preprocessing Techniques using Space and NLTK
- Word Tokens
- StopWord Removal
- Count Vectorizer
- Tf-Idf Vectorizer