Data Science & Machine Learning

Pre-Requisites

- Basic Python (Intermediate Level)
- Git

Mathematics for Data Science

- Statistics – Descriptive & Inferential Stats
- Probability – Basic & Conditional Probability, Bays Theorem
- Algebra Mathematics – Linear & Polynomial Equations
- Matrices & Vectors – properties, operations, use cases
- Calculus

Data Collection Tools & Techniques

- Web Scrapping
- Logging
- DBMS Queries
- Data Using APIs

Data Engineering or Data Preprocessing

- Data Formats & Structures (csv, tsv, excel sheet etc.)
- Assessing Data for Quality Cheque
- Data Wrangling Techniques
    - Discovering
    - Structuring
    - Cleaning
    - Enriching
    - Validating
    - Publishing
- Data Transformations
    - One Hot Encoding
    - Label Encoding
    - Normalization of Data

Data Analysis or Getting Insights from Data

- Process of Question Building
- Answering Common Questions related to data
- Application of Stats to find useful information out of data
- Data Modeling
- Hypothesis Testing

- Text Analysis

Data Visualization

- Aesthetics of Plots
- Creating Charts, Plots and Maps
- Bar Chart, Box Plot, Histograms, Line Plot
- Scatter Plot, Violin plots, Word Cloud, Maps etc.
- Exploratory Data Analysis
- Understanding Trends, Outliers, and Pattern in Data
- Creating Live Plots

Machine learning

- History, Scope and Future of Machine Learning
- Supervise Machine Learning Techniques
    - Regression Algorithms
        - SLR, Multiple Linear Regression, Multivariant Linear Regression
        - Polynomial Regression
        - Lasso & Ridge Regression
        - Gradient Descent
    - Classification Algorithms
        - Logistic Regression
        - Decision Trees Classifiers
        - Random Forest Classifiers
        - Naïve Bays Classifiers
        - K-Nearest Neighbors (KNN)
        - Support Vector Machines (SVM)
- Unsupervised Machine learning Techniques
    - Clustering
        - K-means Clustering Algorithm
        - DBSCAN Clustering Algorithm
    - Dimensionality Reduction
        - Linear Discriminant Analysis
        - Principle Component Analysis
- Text Processing
    - Bag of Words, TF, IDF
    - Sentiment Analysis
    - Word Clouds
- Evaluation Matrices
    - Mean Square, Mean Absolute, RSS and TSS errors
    - R2 Score for Regression Accuracy
    - ROC and AUC Curves for Performance Measuring

    - Precision & Recall Matrix
    - Accuracy Score for Classification Accuracy
- Optimization Techniques
    - Hyper Parameter Tuning
    - Grid Search
    - Cross Validation
    - Early Stopping
- Introduction to Deep Learning
    - Neural Networks

spark

g Data

istributed computing

istributed Storage

ata Analysis on Bigdata using Pyspark

lachine Learning on Bigdata using Pyspark

se Studies

arious Case Studies Related to Data Science & Machine Learning

pstone Project

or Each Section and Algorithm, we will Create a Capstone Project which will show case

ur Detailed Knowledg

Python Modules Used in This Course

- Exploratory Data Analysis
    - numpy, scipy, pandas, matplolib, seaborn, plotly, folium
- Data Scrapping
    - Beautiful soup, Requests
- Machine Learning
    - Sklearn, tensorflow, keras