

Data Science & Machine Learning

Pre-Requisites

- Basic Python (Intermediate Level)
- Git
- Basic Linux

Mathematics for Data Science

- Statistics Descriptive & Inferential Stats
- Probability- Basic & Condition Probability, Bayes' theorem
- Algebra Mathematics Linear & Polynomial Equations
- Matrices & Vectors properties, operation, use Cases
- Calculus

Data Collection Tools & Techniques

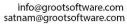
- Web Scraping
 - Logging
 - **DBMS** Queries
 - **Data Using APIs**

Data Engineering or Data Preprocessing

- Data Formats & Structures (csv, tsv , excel sheet etc.)
- Assessing Data for Quality Cheque























- **Data Wrangling Techniques**
 - Ø Discovering
 - Ø Structuring
 - Ø Cleaning
 - Ø Enriching
 - Ø Validating
 - Ø Publishing

Data Transformation . One Hot Encoding . Label Encoding . Normalization of Data

Data Analysis or Getting Insights from Data

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

Data Visualization

- Aesthetics of plots
- Creating charts, plots and maps
- O Bar Chart , Box plot, Histograms , line plot
- O Scatter plot, Violin plots ,Word Cloud, Maps etc.
- Exploratory Data Analysis
- O Understanding Trends ,Outliers ,and pattern in Data
- Creating Live Plots

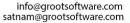
Machine learning

- History ,Scope and Future of Machine Learning
- Supervised Machine Learning Techniques
 - Regression Algorithms

1.SLR , Multiple Linear Regression , Multivariate Linear Regression

























- 2.Polynomial Regression
- 3.Lasso & Ridge Regression
- 4. Gradient Descent

Ø Classification Algorithms

- 1.Logistic Regression
- 2. Decision Trees Classifiers
- 3. Random Forest Classifiers
- 4. Naive Bayes Classifiers
- 5.k-Nearest Neighbors (KNN)
- 6.Support Vector Machines(SVM)

v Unsupervised Machine learning Techniques

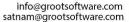
- Hypothesis Testing
- Predictive Analysis
- Text Analysis
- Clustering
- K- means clustering Algorithm
- DBSCAN clustering Algorithm ,Dimensionality Reduction
- Linear Discriminant Analysis
- Principal component Analysis
 - Means Square, Mean Absolute ,RSS and TSS errors
 - R2 Score for Regression Accuracy
 - ROC and AUC Curves for performance Measuring
 - · Classification Reports & Confusion Measuring
 - · Precision & Recall Matrix
 - Accuracy Score for classification Accuracy
- v Optimization Techniques
- v Hyper Parameter Tuning
- v Grid Search
- v Cross Validation
- v Early Stopping

Pyspark

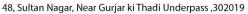
- v Big Data
- v Distributed computing

























- v Distributed Storage
- v Data Analysis on Bigda using Pyspark
- v Machine Learning on Bigdata using Pyspark

Case Studies

v Various Case Studies Related to Data Science & Machine Learning

Capstone Project

For Each Section and Algorithm, we will Create a Capstone Project which will show case your Detailed Knowledge

Python Modules Used in This Course

Exploratory Data Analysis

 \circ Text Processing \circ Bag of Words, TF, IDF \circ Sentiment Analysis \circ Word Clouds

Evaluation matrices

Numpy ,scipy pandas matplotlib seaborn ,plotly ,folium

Data Scraping ,Beautiful soup ,Requests

Machine learning ,Sklearn, tensorflow,keras





