

91116 21112  
8008884446

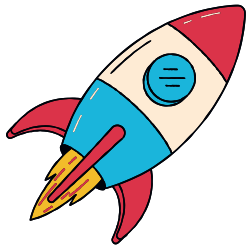
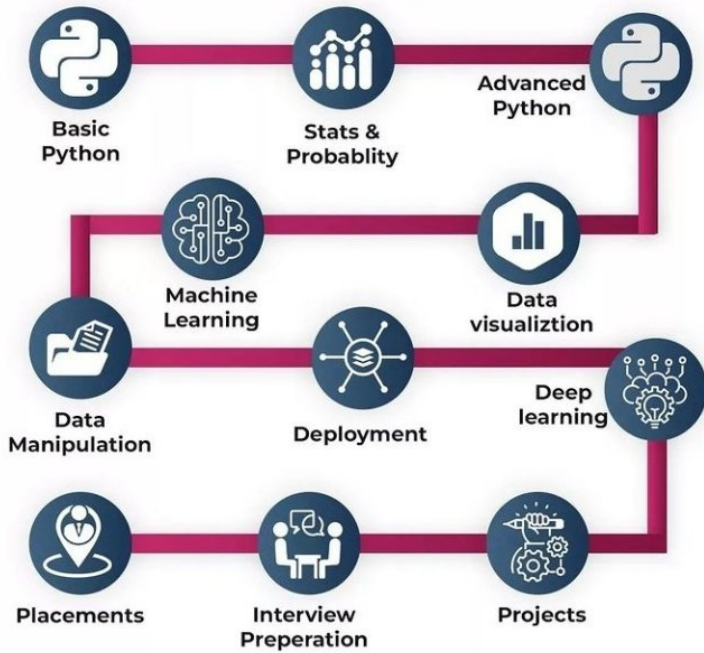


**dataminds**  
TRAINING | PLACEMENTS | CONSULTING

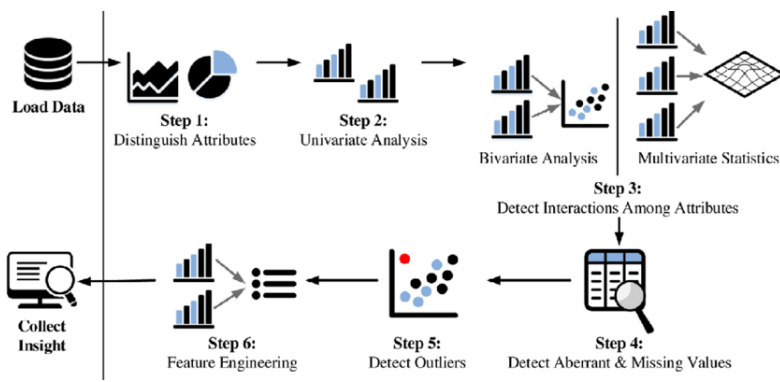


**Online Classroom**

# DATA SCIENCE ROADMAP



**NUMPY  
PANDAS  
MATPLOTLIB  
SCI-KIT LEARN**



- ➔ **Data Science Essentials**      **2 Months**
- ➔ **Data Science & AI**              **3 Months**
- ➔ **Job Oriented Special Programme**

## PYTHON PROGRAMMING

- What is Python?
- History
- Why is Python preferred for Data Science?
- Installation of Python/Jupyter Notebook/ SPYDER

### Functions

- User defined functions
- Parameters
- Nested functions
- Local and Global variables
- Alternate Keys
- Lambda functions
- Sorting Lists and Dictionaries
- Sorting Collections

### Basic of Python:

- Keywords
- Built-in functions
- String Formatting
- Lists
- Loops
- Tuples
- Indexing
- Slicing
- Sequences
- Dictionaries
- Sets
- Importing and exporting data from python into various formats

### OOPS

- Methods and Inheritance
- Abstraction and Encapsulation
- Classes
- Walking Directory Trees
- Initializes
- Instance methods
- Class methods
- Data Static Methods
- Expressions
- Module Aliases
- Math functions
- Random Numbers
- Package Installation Methods and other libraries
- Introduction to Numpy, Pandas
- Plotting in Python
- Creating Data Frames
- Data Manipulation
- Slicing and Dicing

### Error and Exception Handling

- Errors in Python
- Abnormal termination
- Exception handling methods
- Ignoring Errors
- Assertions and effective usage of assertions

## INTRODUCTION TO DATA SCIENCE

- What Is Data Science?
- What is Big Data?
- What is Machine Learning?
- What is Analytics?
- What is Data analysis and Data Mining?
- Analytics project life cycle
- Real life applications, projects and career paths of Data Science

## STATISTICS

- Definition and computation of probability
- Sampling and Sampling distributions
- Prediction and Confidence Interval- Computation and Analysis
- Measurement of central tendencies and it's applications
- Spreads, Distributions (Normal, Z-distribution, Binomial, Poisson) and various types of probability distributions (Continuous and discrete)
- Measures of shape (Skewness and Kurtosis)
- Measures of relationship between variables(Correlation, causation)
- Hypothesis Testing (t-test, Chi-square, Anova)
- Measures of Dispersion ( Variance, Std. deviation,Range)
- Missing Value theorem
- Contingency Table

## Exploratory Data Analysis (EDA) and Data Visualization

- What is EDA and why is it required?
- Outlier treatment
- Data distributions and transformations
- Graphs
- Bar charts
- Histogram
- Box-Whisker plot
- Scatter plot
- Variable selection

## Machine Learning

### Crisp-DM Process Supervised Learning:

- What is supervised learning
- Algorithms in Supervised learning
- Steps in Supervised learning

### Regression & Classification:

- Regression vs classification
- Computation of co-relation coefficient and Analysis
- Performance and accuracy measurement of a Model
  - Naive Baye's classifier
- Model Training, Validation and Testing
- Ordinary Least squares
- Variable selection
- R-Square coefficient and RMSE as a strength of model
- Prediction and confidence interval
- determination and application
- Proviso of Regression
- Dummy variables
- Types of Regression: Linear and Logistic (Sample and multiple)
- Sum of least squares
- ROC and AUG curves
- Homoscedasticity and Heteroscedasticity
- Mutlicollinearly and vift
- Confusion matrix
- Techniques to improve accuracy and performance of regression models

### Decision Trees & Randon Forest Test

- Introduction to Decision tree Algorithms and it's applications
- Classification and regression trees-CART models, ID3, C4.5
- CHAID analysis
- Building Decision Trees using R
- Decision nodes and leaf nodes
- Variable Selection, Parent and child nodes branching
- Stopping Criterion
- Tree pruning
- Depth of a tree
- Ovalfitting
- Metrics for decision trees-Gini impurity, Information Gain, Variance Reduction Regression using desires tree
- Interpretation of a decision tree using if-else
- Pros and cons of a decision tree
- Introduction to Randon forest test and it's applications
- Why Randon forest test?
- Tree bagging
- Models and algorithms in Randon Forest test
- Training Data set, Tree grouping and decision making on majority voting
- Boosting algorithms-Gradient Boosting, Adaptive boosting- Adabocet Xgboost ( Advanced)
- Accuracy estimation using cross validation

## Data Visualization Using Tableau

### Introduction

- Installation of Trial Version of Tableau Public
- Design Flow
- Data Visualization
- Connecting Tableau to various Data Sources
- Measures and Dimensions
- Colors, Labeling and formatting
- Exporting Work sheet

### Advance concept of Tableau

- Trend Line Analysis
- Dash Board Creation
- Formatting in tableau
- Forecasting using Exponential Smoothing
- Granularity and Trimming
- Seasonality
- Animations

### Basics of Tableau

- A-B Ad-hoc Testing
- Aliases
- Reference Line
- Anomaly detection
- Sorts and Filters
- Time Series
- Chart plotting
- Heat Maps
- Data Joining
- Data Blending

### KNN-Algorithm

- What is KNN and wtry do we WAY it?
- KNN-algaithm and regression
- Curse of dimensionality and brief Introduction to dimension reduction
- KNNI-outer treatment and anomaly detection
- Cross Vadation
- Pros and cons of KNN

### Clustering

- Introduction to clustering
- Real-fle applications of clustering
- Distance measurement metriods
- Hierarchical clusteting
- K-Means clustening and skew plot
- Assignment

### Unsupervised Learning:

- What is unsupervised learning?
- Algorithms in unsupervised learning
- Steps in unsupervised learning

### Text Mining

- Introduction to Text Mining
- Applications
- Structured and unstructured data
- Extracting unstructured text from files and websites
- Data cleaning and reshaping
- Terminologies in Teal Mining
- Text clustering and categorization
- Word cloud
- N-gram charts
- Sentiment Analysis
- Twitter Analytics
- Natural Language processing

### Dimensionality Reduction:

- Introduction to dimensionality reduction and it's necessity
- Principal Component Analysis (PCA)
- Singular Value Decomposition (SVD)
- Kernel-PCA
- Linear Discriminate Analysis
- Feature extraction
- Advantages and applications of Dimensionality reduction

### Forecasting\*

- Introduction to forecasting
- Applications
- Data Manipulation and Cleaning
- Time Series
- Time Series forecasting
- Components of Time Series- Trend, Seasonality, Randomness
- Trend Analysis
- Forecasting methods
- Smoothing Methods
- Modeling Random Components
- Modeling for stationary time series
- ETS Model
- Auto regressive Model
- Moving Average Model
- ARIMA Model
- ETS Model
- Anomaly Detection
- Transformations
- Growth curve
- ARCH & GARCH Models

### Association rules:

- Introduction
- Importance of Association rules
- Metrics of rules-Lift, Support, Confidence, Conviction
- Apriority Model
- Market Basket Analysis
- Algorithm implementation and tuning
- Applications

### Deep Learning\*

#### Introduction to Neural Networks\*

- Introduction to NN
- Basic understanding of ANN
- Basic understanding of CNN
- Basic understanding of RNN
- Basic understanding of Back propagation
- Basic understanding of Feed Forward

Computer Vision  
Natural Language  
Proessing

**LIVE** Projects

Interview Faq's  
Cheat Sheets

Tool Kits

CV Preparation

Linkedin Optimization

Applicant Tracking System

DATA MINDS ANALYTICS

#401, GREEN HOUSE, BESIDE ADITYA TRADE CENTRE , AMEERPET