Phase 1: Introduction to Data Science and AI & ML

- Data Science, AI & ML
- Use Cases in Business and Scope
- CRISP
  Modeling

Phase 2: Python Essentials (Tutorial)
Programming

- Commands and Syntax
- Packages and Libraries
- Introduction to Data Types
- Data Structures in Python - Set, Tuple, Dictionary, Lists, Data Frames
- Importing and Exporting Data.
- Control structures and Functions

Descriptive Statistics

- Data exploration (histograms, bar chart, box plot, line graph, scatter plot)
- Qualitative and Quantitative Data
- Measure of Central Tendency (Mean, Median and Mode),
- Measure of Positions (Quartiles, Deciles, Percentiles and Quantiles),
- Measure of Dispersion (Range, Median, Deviation, Variance and Standard Deviation)
- Other Measures: Quartile and Percentile, Interquartile Range

Phase 3: Statistical Analysis
Initial Data Analysis

- Relationship between attributes: Covariance, Correlation Coefficient, Chi Square
- Measure of Distribution (Skewness / Kurtosis), Box-Whisker Plot and statistical graphs

Probability

- Probability (Joint, marginal and conditional probabilities)
- Probability distributions (Continuous and Discrete)
- Density Functions and Cumulative functions

Phase 4: Data Acquisition

- Gather information from different sources.
- Internal systems and External systems.
- Web APIs, Open Data Sources, Data APIs, Web Scrapping
- Relational Database access (queries) to process/access data.

Phase 5: Data Pre-processing and Preparation

- Data Munging, Wrangling

Phase 6: Data Quality and Transformation

- Data imputation
- Data Transformation (minmax, log transform, z-score transform etc.,).
- Binning, Classing and Standardization.
- Outlier/Noise & Anomalies

Phase 7: Handling Text Data

- Bag-of-words
- Regular Expressions
- Sentence Splitting and Tokenization
- Punctuations and Stop words, Incorrect spellings
- Properties of words and Word cloud
- Lemmatization and Term-Document computation
- Sentiment Analysis (Case Study)

Phase 8: Principles of Big Data

- Introduction to Big Data
- Challenges of processing Big Data (Volume, Velocity and Variety perspective)
- Use Cases

Phase 9: Big Data Frameworks – Hadoop, Spark and NoSQL

- Hadoop: In-depth understanding
- Essential Algorithm (Word count, Page Rank)
- Spark: RDDs, Streaming and Spark ML
- NoSQL concepts (CAP, ACID, NoSQL types)

Phase 10: Data Visualization

- Science of Visualization
- Visualization Periodic Table
- Aesthetics and Storytelling
- Concepts of measurement
- Concepts of Size, Shape, Color
- Various Visualization types
- Bubble charts/ Gauge charts
- Geo-maps (Chlorpeths)
- Tree map / Heat map

Phase 11: Sampling and Estimation

- Sample versus population
- Sample techniques (simple, stratified, clustered, random)
- Sampling Distributions
- Parameter Estimation
- Unbalanced data treatment

Phase 12: Inferential Statistics

- Develop an intuition how to understand the data, attributes, distributions
- Procedure for statistical testing, etc.
- Test of Hypothesis (Concept of Hypothesis testing, Null and Alternative Hypothesis)
- Cross Tabulations (Contingency table and their use, Chi-Square test, Fisher)
- Samples t test
- One-way ANOVA (Post hoc tests: Fisher's LSD, Tukey's HSD).
- Z-test and F-test.

Phase 13: Linear Regression

- Regression basics: Relationship between attributes (Covariance and Correlation)
- Relationship between multiple variables: (Linear, Multivariate) in prediction.
- Residual Analysis
- Identifying significant features, feature reduction using AIC, multi-collinearity
- Heteroscedasticity
- Hypothesis testing of Regression
- Confidence intervals of Slope
- R-square and goodness of fit
- Influential Observations – Leverage

Phase 14: Multiple Linear Regression

- Polynomial Regression
- Regularization methods
- Lasso, Ridge and Elastic nets
- Categorical Variables in Regression

Phase15:Logistic Regression

- Error measures (ROCR)
- Logistic Regression in classification

Unit 16: Forecasting models

- Trend analysis
- Cyclical and Seasonal analysis
- Smoothing; Moving averages; Box-Jenkins, Holt-winters, Auto-correlation,
- ARIMA Examples: Applications of Time Series in financial markets

Phase 17: Foundations for ML

- ML Techniques overview
- Validation Techniques (Cross-Validations)
- Feature Reduction/Dimensionality reduction
- Principal components analysis (Eigen values, Eigen vectors, Orthogonality)

Phase 18: Clustering

- Distance measures
- Different clustering methods (Distance, Density, Hierarchical)
- Iterative distance-based clustering;
- Dealing with continuous, categorical values in K-Means
- Constructing a hierarchical cluster
- K-Medoids, k-Mode and density-based clustering
- Measures of quality of clustering

Phase 19: Classification

Naïve Bayes Classifier
- Model Assumptions, Probability estimation
- Required data processing
- M-estimates, Feature selection: Mutual information
- Classifier

K-Nearest Neighbors
- Computational geometry; Voronoi Diagrams; Delaunay Triangulations
- K-Nearest Neighbor algorithm; Wilson editing and triangulations
- Aspects to consider while designing K-Nearest Neighbor

Support Vector Machines
- Linear learning machines and Kernel space, Making Kernels and working in feature space
- SVM for classification and regression problems.

Decision Trees
- Detailed understanding of the Tree formation and algorithms applied

Ensembles methods
- Bagging & boosting and its impact on bias and variance
- Boosting Algorithms
- Random forest
- Gradient Boosting Machines and XGBoost

Phase 20: Association Rule mining

- The applications of Association Rule Mining: Market Basket, Recommendation Engines, etc.
- A mathematical model for association analysis; Large item sets; Association Rules
- Application examples: Association analysis vs. classification

Phase 21: Foundations for AI

- AI: Application areas
- AI Basics (Divide and Conquer, Greedy, Branch and Bound, Gradient Descent)
- NN basics (Perceptron and MLP, FFN, Backpropagation)

Phase 22: Convolution Neural Networks

- Image classification
- Text classification
- Image classification and hyper-parameter tuning
- Emerging NN architectures

Phase 23: Recurrent Neural Networks

- Building recurrent NN
- Long Short-Term Memory
- Time Series Forecasting

Phase 24: Deep Learning

- Auto-encoders and unsupervised learning
- Stacked auto-encoders and semi-supervised learning
- Regularization - Dropout and Batch-Norm

Phase 25: Project
- Case Study : As per the learner's choice