

Course Content: AI and Machine Learning

Lesson 01: Introduction to Data Science and AI & ML

- Data Science, AI & ML
- Use Cases in Business and Scope
- Scientific Method
- Modeling Concepts
- CRISP-DM Method

Lesson 02: Python Essentials Programming

- Commands and Syntax
- Packages and Libraries
- Introduction to Data Types
- Vectors, Matrices, Arrays, Lists, Factors, Data Frames
- Importing and Exporting Data.
- Control structures and Functions

Lesson 03: 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, Absolute deviation about median, Variance and Standard deviation), Anscombe's quartet
- Other Measures: Quartile and Percentile, Interquartile Range

Lesson 04: Statistical Analysis

- Relationship between attributes: Covariance, Correlation Coefficient, Chi Square
- Measure of Distribution (Skewness and Kurtosis), Box and Whisker Plot (Box Plot and its parts Using Box Plots to compare distribution) and other statistical graphs.

Lesson 05: Probability

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

Lesson 06: 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.

Lesson 07: Data Pre-processing and Preparation

- Data Munging, Wrangling
- Plyr packages
- Cast/Melt

Lesson 08: Data Pre-processing and Preparation

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

Lesson 09: Data Visualization

- Science of Visualization
- Visualization Periodic Table
- Aesthetics and Story telling
- Concepts of measurement - scales of measurement
- Design of data collection formats with illustration
- Principles of data visualization - different methods
- of presenting data in business analytics.
- Concepts of Size, Shape, Color
- Various Visualization types
- Bubble charts
- Geo-maps (Chlorpeths)
- Gauge charts
- Tree map
- Heat map
- Motion charts
- Force Directed Charts etc.,

Lesson 10: Sampling and Estimation

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

Lesson 11: Linear Regression

- Regression basics: Relationship between attributes using Covariance and Correlation
- Relationship between multiple variables: Regression (Linear, Multivariate) in prediction.
- Residual Analysis
- Identifying significant features, feature reduction using AIC, multi-collinearity

- Non-normality and Heteroscedasticity
- Hypothesis testing of Regression Model
- Confidence intervals of Slope
- R-square and goodness of fit
- Influential Observations – Leverage

Lesson 12: Multiple Linear Regression

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

Lesson 13: Non-Linear Regression

- Logit function and interpretation
- Types of error measures (ROCR)
- Logistic Regression in classification

Lesson 14: 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

Lesson 15: 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

Lesson 16: Naïve Bayes Classifier

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

Lesson 17: 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

Lesson 18: Support Vector Machines

- Linear learning machines and Kernel space, Making Kernels and working in

feature space

- SVM for classification and regression problems.

Lesson 19: Decision Trees

- ID4, C4.5, CART

Lesson 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
- Apriori: Constructs large item sets with mini sup by iterations; Interestingness of discovered association
- Application examples; Association analysis vs. classification
- FP-trees

Lesson 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)

Lesson 22: Convolution Neural Networks

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

Lesson 23: Recurrent Neural Networks

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

Lesson 24: Deep Learning

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