SIEVE *SOFTWARE*

Flat. No: 307A, 3rd floor, adhitya enclave, nilgiri block, ameerpet, hyd. 8019242423, 7386977110...

Data Science

Data Science with R – Data Analytics – Part I

- Programming Foundation in R
  - Data object: Vectors, Matrices, Data Frames, and Lists
  - Common functions
  - R - Studio Environment and package management
  - Local data input/output
  - Introduction to R data visualization
  - Data sorting and merging
  - String manipulation
  - Dates and times
  - Connecting to an external database

Data Science with R – Data Analytics – Part II

- Data Manipulation with 'dplyr'
  - Join
  - Subset
  - Advanced manipulations with dplyr
- Data Visualization with 'ggplot2'
  - Histogram
  - Point graphics
  - Columnar graphics o Line charts
  - Pie charts o Box plots
  - Scatter plots
  - Visualizing multivariate data
  - Matrix - based visualizations
  - Maps

Data Science with Python – Data Analytics – Part-I

- Python Programming Language I
  - Simple Values and Expressions
  - Functions
  - Lists
  - Conditions
  - Functional Programming: map, filter, and reduce
- Python Programming Language II
  - String operations
  - File input/output and searching
  - Data Structures:
    - ✓ Mutating operations on Lists, Tuples, Sets and Dictionaries
- Python Programming Language III
  - Control flows
  - Errors and exceptions
  - Object Oriented Programming

Data Science with Python – Data Analytics – Part-II

- Numpy and Scipy
  - Basic data structure and operations
  - Matrices and linear algebra
  - Stats module
  - Random Sampling
- Pandas
  - Series and data frame

- I/O of pandas data frame
- Concatenation and merge
- Arithmetic, drop, apply and describe
- Selection and filter
- Missing values
- Grouping and aggregation
- Time series
- Interacting with data base

- Matplotlib and Seaborn
  - Basic plots
  - Statistical plots:
    - ✓ Scatter plots
    - ✓ Histogram
    - ✓ Boxplot
    - ✓ Barchart
  - Multiple figures
  - Advanced plots with seaborn
- Python lab : Linear Regression from scratch

Data Science with R –statistics – Part-I

- Foundations on Statistics
  - Descriptive Statistics
    - ✓ Measures of Centrality
    - ✓ Measures of Variability
      - ➢ Frequency, Proportion & Contingency Tables
      - ➢ Continuous probability distribution
        - ∓ Standard normal distribution
        - ∓ F-distribution
        - ∓ Student t-distribution
        - ∓ Chi-square distribution
      - ➢ Discrete probability distribution
        - ∓ Binomial distribution
        - ∓ Negative binomial distribution
        - ∓ Poisson distribution
      - ➢ Computing probability with Normal Distribution
      - ➢ Central limit theorem for sampling variations
    - ✓ Confidence interval-computation and analysis
    - ✓ Correlation
    - ✓ Hypothesis Testing
      - ➢ Parametric Tests
        - ∓ One Sample, 2 sample -test
        - ∓ 1 sample Z-test
        - ∓ Paired t test
        - ∓ One proportion, 2 proportion test
        - ∓ F-test
        - ∓ One--way ANOVA
        - ∓ Chi- Test of Independence
      - ➢ Non-Parametric Tests
        - ∓ 1 sample sign test
        - ∓ Mannwitney test
        - ∓ Kruskal wallis test
        - ∓ Mood's Median test
- Introduction to Machine Learning
  - ✓ Supervised Learning
    - ➢ Regression

- Classification
- Unsupervised Learning
  - Clustering
  - Dimension Reduction
- Amusingness & Imputation
  - Types of Amusingness
    - MCAR
    - MAR
    - MNAR
  - Basic Methods of Imputation
    - Mean Value Imputation
    - Simple Random Imputation
    - Regression Prediction
  - K-Nearest Neighbors
    - Voronoi Tessellations
    - KNN for Classification
    - KNN for Regression
    - Distance Measures
- Linear Regression I
  - Simple Linear Regression
    - From a Mathematical Standpoint
    - Accuracy of the Coefficient Estimates
    - Performing Hypothesis Tests
    - Constructing Confidence Intervals
  - Assumptions & Diagnostics
  - Transformations
    - Power Transformation
    - Box-Cox Transformation
  - The Coefficient of Determination R^2
- Linear Regression II
  - Multiple Linear Regression
    - From a Mathematical Standpoint
  - Assumptions & Diagnostics
  - Potential Problems
  - Research Questions
  - Variable Selection
  - Factors
  - Interactions
  - Higher--Order Terms

Data Science with R –Machine Learning – Part-II

- Generalized Linear Models o Logistic Regression
- The Curse of Dimensionality
  - Ridge Regression
  - Lasso Regression
  - Cross--Validation
  - Bias/Variance Tradeoff
  - Density
  - Principal Component Analysis
- The Curse of Dimensionality
  - Density
  - Principal Components Analysis

Data Science with R –Machine Learning – Part-III

- Classification
  - Feature Selection
  - Support Vector Machines

- ✓ Decision Trees
- ✓ Pruning/Purity/Entropy/GINI
- ✓ Random Forests
- ✓ Bagging
- ✓ Boosting

- Cluster Analysis
  - ✓ K -Means Clustering
  - ✓ Agglomerative Clustering
  - ✓ Hierarchical Clustering

Data Science with R –Machine Learning – Part-IV

- Association Rules
  - ✓ Market Basket Analysis
- Naive Bayes Analysis
- Introduction to Natural Language Processing
  - ✓ Creating corpus: stemming and lemmatization
  - ✓ POS tag and chunking
  - ✓ Text classification
- Time Series Analysis
  - ✓ Smoothing
  - ✓ Seasonal Decomposition
- ARIMA

Data Science with Python –Machine Learning

- Machine Learning Recap / Linear Regression
  - Introduction to scikit learn
  - Simple linear regression
  - Multiple linear regression
  - Stats module
- Classification part I
  - Logistic regression
  - Discriminate analysis
  - Naïve Bayes
- Model Selection
  - Cross--validation
  - Bootstrap
  - Feature selection
  - Regularization
  - Grid search
- Classification part II
  - Support vector machine
  - Decision tree
  - Random forest
- Unsupervised learning
  - Principal Components Analysis
  - K - Means and Hierarchical Clustering