

Data-science - R

Duration: 3 Days

INTRODUCTION

- Data science & its importance
- Key Elements of Data Science
- Introduction to ML
- Artificial Intelligence & Machine Learning Introduction
- Who uses AI?
- AI for Banking & Finance, Manufacturing, Healthcare, Retail and Supply Chain
- Supervised & Unsupervised Learning
- Regression & Classification Problems
- What makes a Machine Learning Expert?
- What to learn to become a Machine Learning Developer?
- Overview of Machine Learning Algorithms

R

- R basics (If-Else, Control Structures, Loops, Functions, Data Types)
- Data structures (Vector, Matrix, Dataframe, Lists)
- Indexing, Data Processing
- Mathematical computing basics
- Getting Started with Dplyr and tidyr
- Data Acquisition (Import & Export)
- Selection and Filtering
- Combining and Merging Data Frames

STATISTICS and EDA

- Introduction to Visualization
- Visualization Importance
- Working with R visualization libraries like ggplot2
- Creating Line Plots, Bar Charts, Pie Charts, Histograms, Scatter Plots
- Understanding Box plots
- Understanding Probability Distributions, Violin Plots
- Correlations and Heatmaps
- Summary Statistics
- Central Tendency measures
- Measures of dispersion
- Normal Distributions and z-score
- Missing Value Imputation
- Outlier Detection and handling
- Advanced EDA techniques

APPLIED MACHINE LEARNING

- Machine Learning Algorithms Generic Concepts
 - Sample and Population
 - Bias-Variance Trade off
 - Overfitting and Underfitting
 - Cross Validation
 - Regularization techniques
 - Hyperparameter tuning & grid search optimization
- Linear Regression
 - Regression Problem Analysis
 - Mathematical modelling of Regression Model
 - Gradient Descent Algorithm
 - Use cases
 - Regression Table
 - Model Specification
 - L1 & L2 Regularization
 - Building simple Univariate Linear Regression Model
 - Multivariate Regression Mode
 - R2, p-value, RMSE and residual plots
- Logistic Regression
 - Assumptions
 - Sigmoid function
 - ROC Curve
 - Model Specification
 - Confusion Matrix
 - Accuracy, Recall, Precision and F1 Score
- Decision Trees
 - Forming a Decision Tree
 - Components of Decision Tree
 - Mathematics of Decision Tree
 - Decision Tree Evaluation
- KNN
 - Components of Decision Tree
 - Mathematics of Decision Tree
 - Metrics for evaluation
- Support Vector Machine
 - Concept and Working Principle
 - Mathematical Modelling
 - Optimization Function Formation
 - The Kernel Method and Nonlinear Hyperplanes
- Ensemble Models
 - Bagging
 - Boosting
 - Stacking
 - Voting Classifier
 - Random Forest
- Unsupervised Machine Learning algorithms
 - Clustering with K-meansClustering
 - Clustering with Hierarchical Clustering
 - Advanced clustering techniques and use cases
 - Dimensionality Reduction
 - PCA
- Text Mining and NLP
 - Sentiment Analysis

- Topic Summarization
- Topic Modelling
- Bag of Words and Tf-IDF
- Cosine Similarity of terms, documents concepts
- Text Cleaning and Preprocessing using Regex
- Tokenization, Stemming and Lemmatization

CASE STUDY AND PROJECTS

Students would be given challenging real-life cases to solve – just to augment their learning skills