Contact Us : +91 910 0036 789
Email : info@vuematech.com
Visit Us : www.vuematech.com

Data Science Course Content

Duration: 60 Days

Section: 01 - Introduction to Data Science

- What is Data Science?
- Data Science Life cycle
- Different types of Data Science tasks

Section: 02 - Introduction to R-Programming

- Importance of R
- R and R-studio installation
- Data types and Data structures
- Arithmetic, logical operations
- Creating new variables or updating existing Variables
- Conditional statements and loops
- String manipulations
- Packages and functions in R
- Data Frame operations
- Getting data into R from flat files
- Data Inspection and Manipulation

Section: 03 - Business Statistics

- Probability refresher
- Descriptive statistics
- Measures of central tendency
- Measures of spread
- Sampling
    - Need for sampling?
    - Different types of sampling
- Sampling Distributions
    - Normal Distribution
    - Characteristics of Normal Distribution
    - Binomial distribution
- Inferential statistics
- Hypothesis testing
    - Type I error
    - Type II error
    - Null and alternate hypothesis
    - Reject or acceptance criterion
- Correlation, covariance, associations, odds ratio

Section: 04 - Exploratory Data Analysis

- Getting data into R
- Cleaning and preparing the data
- Data conversions
- Handling missing values
- Outlier Detection and Management
- Binning and Data normalization
- Cross Validation
- Dimensionality Reduction

Section: 05 - Data Visualization

- Visualization in R using ggplot2 (plots and charts)
- Histogram
- Bar chart
- Box plot
- Scatter plot
- Correlation
- Spurious correlation
- Correlation vs. Causation
- Visualization using Tableau (Introduction)

Section: 06 - Supervised Learning

- Introduction
- Steps in Supervised Learning
- Regression and Classification
- Training, Validation and Testing
- R-square, RMSE for Regression
- Confusion Matrix
- F-1 Score, Accuracy, Precision and Recall
- Sensitivity and Specificity
- ROC and AUC

Section: 07 - Linear Regression

- Simple Linear Regression
- Cost Functions
- Sum of Least Squares
- Gradient Descent Approach
- Variable Selection
- Model Development and interpretation
- Model Validation and Diagnostics
- Advantages and Disadvantages

Section: 08 - Logistic Regression

- Need for Logistic Regression
- LOGIT link function
- Variable Selection Methods
- Model Development and interpretation
- Measurements of Accuracy
- Model Validation

- Advantages and Disadvantages

Section: 09 - Decision Trees

- Classification and Regression Trees
- Process of Tree building
- Choosing variables for Decision nodes
- Measures of Impurity
- Entropy, GINI Index and Information Gain
- Over fitting and Pruning
- Advantages and Disadvantages

Section: 10 - Re-Sampling and Ensembles Methods

- Bagging
- Random Forests
- Boosting
    - Gradient Boosting Machines (GBM)
    - Extreme Gradient Boosting-XG Boost

Section: 11 - KNN (K-nearest Neighbors')

- Selection of K
- Normalization of variable
- Model Development and Validation
- Advantages and Disadvantages of KNN

Section: 12 - Support Vector Machines

- Kernels-Linear and Non Linear
- Support Vector Classifier

Section: 13 - Probabilistic Methods (Naïve Bayes)

- Conditional Probabilities
- Classification using Naïve Bayes Approach

Section: 14 - Neural Networks

- Network Topology
- Feed Forward and Backward Propagation models

Section: 15 - Text Mining

- Introduction to Natural Language Processing
- Corpus
- Cleaning Text Data
- Tokenization
- Word Cloud
- Term Document Matrix
- Sentiment Analysis
- Text Classification

Section: 16 - Unsupervised Learning

- Clustering (Segmentation)
- Hierarchical Clustering
- K-Means Clustering
- Distance Measures

Section: 17 - Dimensionality Reduction Techniques

- Principal Component Analysis (PCA)
- Singular Value Decomposition (SVD)

Section: 18 - Time Series Analysis-Forecasting

- Components of Time Series
- Moving Averages
- Exponential Smoothing
- Trend, Seasonality, Randomness
- ARIMA, ARIMAX

Section: 19 - Association Rules (Market Basket Analysis)

- APRIORI

Section: 20 - Recommender Systems

- Collaborative filtering
- User Based Collaborative filtering
- Item Based Collaborative filtering

Section: 21 - Introduction to Python for Data Science

- Python Introduction
- Python installation
- Comparison with R
- Data types and Data structures
- Functions
- Numpy, Pandas, Data frames
- Matplotlib,
- Scikit –Learn
- Use Cases

Section: 22 - Introduction to Big Data Analytics

- Hadoop: Distributed File System
- Map reduce, Hive, Pig

Section: 23 - Spark

- Introduction
- Spark Frame Work
- RDD
- Spark
- PySpark

Section: 24 - Tensor Flow

- Introduction to Tensor Flow
- Introduction Deep Learning with Tensor Flow