Data Science -Machine Learning-Artificial Intelligence

Module 1 - Data Science Introduction
- Course Introduction
- Environment Set-Up
- Jupyter Notebook Overview

Module 2 - Introduction To Basic Statistics Using R And Python

Data Types

- Measure Of central tendency
- Measures of Dispersion
- Graphical Techniques
- Skewness & Kurtosis
- Box Plot
- R
- R Studio
- Descriptive Stats in R
- Python (Installation and basic commands) and Libraries
- Jupyter note book
- Set up Github
- Descriptive Stats in Python
- Pandas and Matplotlib / Seaborn

Module 3 - Probability And Hypothesis Testing

- Random Variable
- Probability
- Probability Distribution
- Normal Distribution
- SND
- Expected Value
- Sampling Funnel
- Sampling Variation
- CLT
- Confidence interval
- Assignments Session-1 (1 hr)
- Introduction to Hypothesis Testing
- Hypothesis Testing with examples
  - ➢ 2 proportion test
  - ➢ 2 sample t test
- Anova and Chisquare case studies

ITS

ITS Data-Solutions Pvt. Ltd.
3rd Floor, Vrajendra Complex, Near Devendra Bus Stop, Naranpura
Ahmedabad 380013 Gujarat India Phone : +91 9879558787, +91 9825306119
www.itsahmedabad.com GST 24AACCI0498K1ZF CIN no : U72200GJ2009PTC056371

Module 4 -Exploratory Data Analysis -1

- Visualization
- Data Cleaning
- Imputation Techniques
- Scatter Plot
- Correlation analysis
- Transformations
- Normalization and Standardization

Module 5 - Linear Regression

- Principles of Regression
- Introduction to Simple Linear Regression
- Multiple Linear Regression

Module 6 - Logistic Regression

- Multiple Logistic Regression
- Confusion matrix
- Receiver operating characteristics curve (ROC curve)

Module 7 - Data Mining Unsupervised Clustering

- Supervised vs Unsupervised learning
- Data Mining Process
- Hierarchical Clustering / Agglomerative Clustering
- Visualization of clustering algorithm using Dendrogram

Module 8 - K-Means

- Non-Hierarchial
- Measurement metrics of clustering - Within Sum of Squares, Between Sum of Squares, Total Sum of Squares
- Choosing the ideal K value using Screen plot / Elbow Curve

Module 9 - Dimension Reduction Techniques

- PCA and tSNE
- Why dimension reduction
- Advantages of PCA
- Calculation of PCA weights
- 2D Visualization using Principal components
- Basics of Matrix algebra

Module 10 - Association Rules

- What is Market Basket / Affinity Analysis
- Measure of association
- Support
- Confidence
- Lift Ratio
- Apriori Algorithm

Module 11 - Recommender System

- User-based collaborative filtering
- Measure of distance / similarity between users
- Driver for recommendation
- Computation reduction techniques
- Search based methods / Item to item collaborative filtering
- Vulnerability of recommender systems

Module 12 - Introduction To Supervised Machine Learning

- Workflow from data to deployment
- Data nuances
- Mindsets of modeling

Module 13 - Decision Tree

- Elements of Classification Tree - Root node, Child Node, Leaf Node, etc.
- Greedy algorithm
- Measure of Entropy
- Attribute selection using Information Gain
- Implementation of Decision tree using C5.0 and Sklearn libraries

Module 14 - Exploratory Data Analysis - 2

- Encoding Methods
  - ➢ OHE
  - ➢ Label Encoders
  - ➢ Outlier detection-Isolation Fores
- Predictive power Score

Module 15: Feature Engineering

- Recurcive Feature Elimination
- PCA

Module 16: Model Validation Methods
- Splitting data into train and test
- Methods of cross validation
- Accuracy methods

Module 17:Ensembled Techniques

- Bagging
- Boosting
- Random Forest
- XGBM
- LGBM

Module 18 - KNN And Support Vector Machines

- Deciding the K value
- Building a KNN model by splitting the data
- Understanding the various generalization and regulation techniques to avoid overfitting and underfitting
- Kernel tricks

Module 19 - Regularization Techniques

- Lasso Regression
- Ridge Regression

Module 20 - Neural Networks

- Artificial Neural Network
- Biological Neuron vs Artificial Neuron
- ANN structure
- Activation function
- Network Topology
- Classification Hyper planes
- Best fit "boundary"
- Gradient Descent
- Stochastic Gradient Descent Intro
- Back Propogation
- Introduction to concepts of CNN

Module 21 - Text Mining
- Sources of data
- Bag of words
- Pre-processing, corpus Document-Term Matrix (DTM) and TDM
- Word Clouds

- Corpus level word clouds
- Sentiment Analysis
  - ➢ Positive Word clouds
  - ➢ Negative word clouds
  - ➢ Unigram, Bigram, Trigram
  - ➢ Vector space Modelling
- Word embedding
- Document Similarity using Cosine similarity
- Extract Tweets from Twitter
- Extract user reviews of the products from Amazon, Snapdeal and TripAdvisor
- Install Libraries from Shell
- Extraction and text analytics in Python

Module 22 - Natural Language Processing

- Sentiment Extraction
- Lexicons and Emotion Mining

Module 23 - Naive Bayes

- Probability – Recap
- Bayes Rule
- Naive Bayes Classifier
- Text Classification using Naive Bayes

Module 24 - Forecasting

- Introduction to time series data
- Steps of forecasting
- Components of time series data
- Scatter plot and Time Plot
- Lag Plot
- ACF - Auto-Correlation Function / Correlogram
- Visualization principles
- Naive forecast methods
- Errors in forecast and its metrics
- Model Based approaches
  - ➢ Linear Model
  - ➢ Exponential Model
  - ➢ Quadratic Model
  - ➢ Additive Seasonality
  - ➢ Multiplicative Seasonality
- Model-Based approaches
- AR (Auto-Regressive) model for errors

- Random walk
- ARMA (Auto-Regressive Moving Average), Order p and q
- ARIMA (Auto-Regressive Integrated Moving Average), Order p, d and q
- Data-driven approach to forecasting
- Smoothing techniques
  - ➢ Moving Average
  - ➢ Simple Exponential Smoothing
  - ➢ Holts / Double Exponential Smoothing
  - ➢ Winters / HoltWinters
- De-seasoning and de-trending
- Forecasting using Python and R

Module 25 - Survival Analysis

- Concept with a business case

Module 26 - End To End Project Description With Deployment

- End to End project Description with deployment using R and Python