INFOSOFT IT SOLUTIONS

Training | Projects | Placements

Revathi Apartments, Ameerpet, 1st Floor, Opposite Annapurna Block,

Infosof solutions, Software Training & Development Institute, +91-9059683947 | +91-9182540872

Business Analytics with R Course

Data Science Project Lifecycle

- Recap of Demo
- Introduction to Types of Analytics
- Project life cycle
- An introduction to our E learning platform

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

Probability And Hypothesis Testing

- Random Variable
- Probability
- Probility 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

Exploratory Data Analysis -1

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

Linear Regression

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

Logistic Regression

- Multiple Logistic Regression
- Confusion matrix
- False Positive, False Negative
- True Positive, True Negative
- Sensitivity, Recall, Specificity, F1 score
- Receiver operating characteristics curve (ROC curve)

Deployment

- R shiny
- Streamlit

Data Mining Unsupervised Clustering

- Supervised vs Unsupervised learning
- Data Mining Process
- Hierarchical Clustering / Agglomerative Clustering
- Measure of distance
- Numeric - Euclidean, Manhattan, Mahalanobis
- Categorical - Binary Euclidean, Simple Matching Coefficient, Jaquard's Coefficient
- Mixed - Gower's General Dissimilarity Coefficient
- Types of Linkages
- Single Linkage / Nearest Neighbour
- Complete Linkage / Farthest Neighbour
- Average Linkage
- Centroid Linkage
- Visualization of clustering algorithm using Dendrogram

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

Association Rules

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

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

Introduction To Supervised Machine Learning

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

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

Exploratory Data Analysis - 2

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

Feature Engineering

- Recurcive Feature Elimination
- PCA

Model Validation Methods

- Splitting data into train and test
- Methods of cross validation
- Accuracy methods

Ensembled Techniques

- Bagging
- Boosting
- Random Forest
- XGBM
- LGBM

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

Regularization Techniques

- Lasso Regression
- Ridge Regression

Neural Networks

- Artificial Neural Network
- Biological Neuron vs Artificial Neuron
- ANN structure
- Activation function
- Network Topology
- Classification Hyperplanes
- Best fit “boundary”
- Gradient Descent
- Stochastic Gradient Descent Intro
- Back Propogation
- Intoduction to concepts of CNN

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

Natural Language Processing

- Sentiment Extraction
- Lexicons and Emotion Mining

Naive Bayes

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

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

Survival Analysis

- Concept with a business case

End To End Project Description With Deployment

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