

# Data Science Machine Learning /AI

## (Python-R-SPSS\* -Excel-Tableau-SAS\* E-Miner )

*Note: \* Commercial Tools are / Subjectively available on Demand. Personal laptop is allowed Office laptops not allowed cause of licence legal privacy*

## Module1: Introduction Data Science

### Part -1 Data science Business Analytics

- Fact of Data Science and Business analytics
- SWOT Analysis of Data Science
- Journey Mathematics-Statistics-Econometrics
- SQL data for Data Science
- NoSQL data for Data Science
- OLTP OLAP for Data information
- Web Application report
- Difference of Machine Meaning AI
- Difference of Data mining Data Science
- DM-CRISP Model processing

## Module 2: Visualization & Summarization

### Part-2: Exploratory Data Analysis

- Data Type
  - Continues
  - Discrete
  - Nominal
  - Ordinal
  - Binary
- Measures of central tendency
  - Mean
  - Median
  - Mode
  - Geomean
  - Harman
  - Trimmed Mean
  - Weighted Mean
  - 95% CI L mean
  - 95% CI U mean
- Data Viability Dispersion
  - Std
  - Variance
  - Coefficient Of Variance
  - Range
  - Min
  - Max
  - skewed
  - kurtosis
  - std Error

- std Skewed
- Error Kurtosis
- IQR
- Five Number Summary
  - Q0 Min
  - Q1 25%
  - Q2 50% median
  - Q3 75%
  - Q4 100%
- Data Visualization & Visual Data Validation
  - Bar chart
  - Pie chart
  - Area plot
  - Scatter plot
  - Surface
  - Stock plot
  - Radar
  - Tree map
  - Waterfall
  - Heatmap
  - Bubble chart
  - Line chart
  - Histogram
  - Standardized plot
  - Stem leaf
  - Boxplot
  - Skewed plot
  - Liptokurtic plot
  - Platonkurtic plot
  - Massokurtic plot
  - PP plot
  - QQ plot

## Part-3: Sampling Techniques Big Data

- Sampling Distributions
- Simple Random
- Skewed Std. Error
- Kurtosis Std. Error
- Central Limit Theorem,
- Sampling from Infinity
- Sampling Distributions for Mean
- Sampling Distributions for proportions

## Part-4: Probability

- Simple Probability
- Marginal Probability
- Joint Probability
- Conditional probability (Bayes' Theorem probability)
- Discrete Distributions
- Binomial Distribution
- Expected Mean
- Variance
- Bivariate distribution
- Covariance
- Hypergeometric Distributions
- Poisson Distribution
- Continuous Distributions

- Random Sample
- Simple Random sample
- Stratified Random sample
- Systematic Random sample
- Cluster random sample

## Module 3: Data Validation Normality

### Part-5: Data Validation Data Normality

- Univariate normality techniques
- Bivariate techniques
- Multivariate techniques
- Q-Q probability plots
- PP plot
- Cumulative frequency
- Steam and leaf analysis
- Histogram Box plot, Z Score test.
- Shapiro-Wilk Test for Normality
- Anderson-Darling Normality

### Part – 6 Data Cleaning outlier treatment

- Outlier treatment with robust measurements
- Outlier treatment with central tendency Mean
- Outlier with Min Max Likelihood methods
- Outlier with Residual Analysis
- Data Imputation with series Central Tendency

### Part-7: Test of Hypothesis

- Null Hypothesis formulation
- Alternative Hypothesis
- One tail Test ,Two tail Test
- One Sample T-TEST
- Paired T-TEST
- Independent Sample T-TEST
- Analysis of Variance (ANOVA),
- ANCOVA
- MANOVA
- Chi-square Pearson
- Kendall Chi-square
- Wald Chi-square
- Kruskal-Wallis Rank Test Chi Square
- Mann-Whitney, Chi Square
- McNemar test Chi Square
- Nagelkerke Chi-square
- Data Transformation

### Part- 8 Data transformation

- Sqrt Transformation
- Log transformation
- Arcsine transformation
- Box- Cox transformation
- Square root transformation
- Inverse transformation
- Min Max Data normalization Rescaling

- PCA Transformation

# Module 4 Machine Learning AI

## Part- 9: Supervised Learning

### 1. Linear Regression (Functional Models)

- Correlation - Pearson, Kendall, Wilcoxon
- SLR Regression
- MLR Regression
- Examination Residual analysis
- Residual QQ plot
- Residual EDA Analysis
- Residual Standardised
- Auto Correlation
- Test of ANOVA Significant
- VIF Analysis
- Test of T-test Significant
- CP Indexing
- Excluding Constant, and excluding constant
- Homoscedasticity
- Heteroscedasticity
- Stepwise regression
- Forward Regression
- Backward Regression
- Multicollinearity
- Cross validation
- MAPE
- Check prediction accuracy
- Standardized regression
- Quadrant Regression
- Transformed Regression
- Dummy Variables Regression

### 2. Logistics Regression (Classification Models)

- Logit regression
- Binary Regression Analysis
- Probit regression
- Ordinal Regression
- Multinomial Regression
- Stepwise Regression
- Backward Regression
- Forward Regression
- Discriminate Regression Analysis
- Multiple Discriminant Analysis
- Test of Associations
- Chi-square strength of association
- Wald Test statistics for Model
- Hosmer Lemshow
- Pseudo R square
- Maximum likelihood estimation
- Model Fit
- Model cross validation
- AIC
- AICC
- BIC (Bayesian information criterion)

### **3. Timeseries (Forecasting Models)**

- 3.1 Navie model
- 3.2 Moving Averages
- 3.3 Weighted Moving Averages
- 3.4 Exponential Smoothing

### **4 Decision Tree**

- 4.1 GINI
- 4.2 Entropy
- 4.3 CHAID
- 4.4 CART
- 4.4 Prunned /Unpruned Tree (Weka)
- 4.5 Random Forestry
- 4.6 Boosting bagging
- 4.7 Ensemble Models

### **5 Naive Bayes**

### **6. KNN**

### **7. SVM**

## **Part-10: Un Supervised Learning**

### **8. PCA/Dimension Reduction Analysis (Un Supervised Learning )**

- Factor Analysis
- Principle component analysis
- Reliability Test
- KMO MSA tests,
- Rotation
- Future Extraction for regression

### **9.Cluster Analysis**

- Hierarchical clustering
- K Means clustering
- Wards Methods,
- Linkage Methods
- Euclidean distance
- Dendogram

## **Part-11: Deep Learning**

### **○ 10 Neural Network**

- 10.1 ANN
- 10.2 CNN
- 10.3 RNN

## **Part-12: Semi Supervised Learning**

- **11 Aprior algorithm**
- **12. Association Mining MBA**
- **13. Recommendation System**

## **Part -13: Model Validation**

- Model Validation and Testing
- Kappa Statistics
- AIC,
- BIC
- Error/Confusion matrices
- ROC

- APE
- MAPE
- Lift Curve,
- Sensitivity
- Misclassification Rating
- Specificity
- Maximum Absolute Error

## Part -14 Text mining

- NLP
- Sentiment Analysis

## Part -15 Model Deployment

- Microsoft Azure
- Google Clod
- Amazon WNS

## Part -16 Big Data and data warehouse architecture

- Data Integration
- ETL transformation
- Data deployment