Day 1: Warm up on Data Science. Skills, technologies, applications

Day 2: Engineering Data Science with R. An introduction

Module 1: Fundamentals of Probability and Statistics

* Introduction to central tendencies
  + Central tendencies
    - Mean
    - Mode
    - Variance
  + Variability and Spread
    - Range
    - Standard deviation and Variance
    - Inter quartile range
  + Outliers
    - Box plots
* Fundamentals of probability
  + Fundamentals
    - Apriori / Theoretical
    - Apostereori / Frequentist
  + Joint probability
  + Marginal probability
  + Conditional probability
  + Bayes theorem
* Probability distributions
  + - Bernoulli, Binomial, Geometric, Poisson
    - Exponential, Normal, t-distributions
  + Confidence levels and confidence intervals
* Inferential Statistics
  + Statistical Hypothesis testing
  + Chi-square test
  + T-test
  + Z-test
  + F-Test
  + Anova

Module 2: Statistics and Probability in Decision Modeling

* Introduction to prediction and classification
  + Linear regression
  + Logistic regression
  + Relationship between variables – predictors and target – in prediction
  + Feature importance, feataure reduction
  + AIC approach
  + Error metrics
  + Interpreting the output – Rsquare, Adjusted R square, degrees of freedom, RMSE, MAPE
* Time Series Analysis
  + Trend Analysis
  + Cyclical and Seasonal Analysis
  + Ssmoothing – Moving average, Auto-correlation, ARIMA
  + VaR
* Additional techniques
  + Bayesian belief nets
  + MLE
  + Naïve Bayes classification

Module 3: Machine Learning Supervised and Unsupervised

* Unsupervised machine learning techniques Clustering
  + Distance measures
  + Iterative distance based clustering
  + Dealing with continuous, categorical values in K-Means
  + Hierarchical cluster
  + Stability check of clusters
* Advanced clustering techniques
  + PCA, SVD, LSA
  + Kernel tricks and kernel functions
* Classification rules
  + Decision trees
* Association Rules
  + Apriori
  + Outlier analysis
  + Decision trees
* KNearest neighbours
* Collaborative filtering
* Matrix factorization
* Neural Networks and applications
  + Perceptron and single layered neural networks
  + Back propagation algorithms
  + Self-organizing maps
  + Auto-encoders
  + Parse auto-encoders
* Support Vector Machines

Module 4: Optimization techniques and decision analysis

Module 5: Artificial Intelligence and Neural Networks

Module 6: Natural Language Processing

Module 7: Data visualization

Module 8: Big Data – the stack and relevance