

## **DATA SCIENCE CURRICULAM**

### **Getting Started With Data Science and Recommender Systems**

- Data Science Overview
- Reasons to use Data Science
- Project Lifecycle
- Data Acquirement
- Evaluation of Input Data
- Transforming Data
- Statistical and analytical methods to work with data
- Machine Learning basics
- Introduction to Recommender systems
- Apache Mahout Overview

### **Reasons to Use, Project Lifecycle**

- What is Data Science?
- What Kind of Problems can you solve?
- Data Science Project Life Cycle
- Data Science-Basic Principles
- Data Acquisition
- Data Collection
- Understanding Data- Attributes in a Data, Different types of Variables
- Build the Variable type Hierarchy
- Two Dimensional Problem
- Co-relation b/w the Variables- explain using Paint Tool
- Outliers, Outlier Treatment
- Boxplot, How to Draw a Boxplot

### **Acquiring Data**

- Discussion on Boxplot- also Explain
- Example to understand variable Distributions
- What is Percentile? Example using Rstudio tool
- How do we identify outliers?
- How do we handle outliers?

- Outlier Treatment: Using Capping/Flooring General Method
- Distribution- What is Normal Distribution?
- Why Normal Distribution is so popular?
- Uniform Distribution
- Skewed Distribution
- Transformation

## Machine Learning In Data Science

- Discussion about Boxplot and Outlier
- Goal: Increase Profits of a Store
- Areas of increasing the efficiency
- Data Request
- Business Problem: To maximize shop Profits
- What are Interlinked variables
- What is Strategy
- Interaction b/w the Variables
- Univariate analysis
- Multivariate analysis
- Bivariate analysis
- Relation b/w Variable
- Standardize Variables
- What is Hypothesis?
- Interpret the Correlation
- Negative Correlation
- Machine Learning

## Statistical And Analytical Methods Dealing With Data, Implementation Of Recommenders Using Apache Mahout And Transforming Data

- Correlation b/w Nominal Variables
- Contingency Table
- What is Expected Value?
- What is Mean?
- How Expected Value is differ from Mean
- Experiment “ Controlled Experiment, Uncontrolled Experiment
- Degree of Freedom
- Dependency b/w Nominal Variable & Continuous Variable
- Linear Regression
- Extrapolation and Interpolation

- Univariate Analysis for Linear Regression
- Building Model for Linear Regression
- Pattern of Data means?
- Data Processing Operation
- What is sampling?
- Sampling Distribution
- Stratified Sampling Technique
- Disproportionate Sampling Technique
- Balanced Allocation-part of Disproportionate Sampling
- Systematic Sampling
- Cluster Sampling
- 2 angels of Data Science-Statistical Learning, Machine Learning

### **Testing and Assessment, Production Deployment and More**

- Multi variable analysis
- linear regration
- Simple linear regration
- Hypothesis testing
- Speculation vs. claim(Query)
- Sample
- Step to test your hypothesis
- performance measure
- Generate null hypothesis
- alternative hypothesis
- Testing the hypothesis
- Threshold value
- Hypothesis testing explanation by example
- Null Hypothesis
- Alternative Hypothesis
- Probability
- Histogram of mean value
- Revisit CHI-SQUARE independence test
- Correlation between Nominal Variable

### **Business Algorithms, Simple Approaches to Prediction, Building Model, Model Deployment**

- Machine Learning
- Importance of Algorithms
- Supervised and Unsupervised Learning

- Various Algorithms on Business
- Simple approaches to Prediction
- Predict Algorithms
- Population data
- sampling
- Disproportionate Sampling
- Steps in Model Building
- Sample the data
- What is K?
- Training Data
- Test Data
- Validation data
- Model Building
- Find the accuracy
- Rules
- Iteration
- Deploy the model
- Linear regression

### **Getting Started With Segmentation of Prediction and Analysis**

- Clustering
- Cluster and Clustering with Example
- Data Points, Grouping Data Points
- Manual Profiling
- Horizontal & Vertical Slicing
- Clustering Algorithm
- Criteria for take into Consideration before doing Clustering
- Graphical Example
- Clustering & Classification: Exclusive Clustering, Overlapping Clustering, Hierarchy

#### **Clustering**

- Simple Approaches to Prediction
- Different types of Distances: 1.Manhattan, 2.Euclidean, 3.Consine Similarity
- Clustering Algorithm in Mahout
- Probabilistic Clustering
- Pattern Learning
- Nearest Neighbor Prediction
- Nearest Neighbor Analysis