**Data Science Course Syllabus**

**INTRODUCTION TO DATA SCIENCE**

The Data Science Overview, Data Science – Why all the excitement? Demand for   
Data Science Professionals, Brief Introduction to Big data and Data Analytics, Life  
cycle of data science, what does Data scientist Do. Tools and Technologies used in data Science.

**STATISTICS**

**FUNDAMENTALS OF MATHEMATICS AND PROBABILITY**

**MACHINE LEARNING**

**INTRODUCTION TO MACHINE LEARNING –**Machine Learning? What is the Challenge? Supervised Learning, Unsupervised Learning, Reinforcement Learning

**LINEAR REGRESSION**– Linear Regression with Multiple Variables, Disadvantage of Linear Models, Interpretation of Model Outputs, Understanding Covariance and Co linearity, Understanding Heteroscedasticity, Case study on Application of Linear Regression for housing price prediction

**LOGISTIC REGRESSION**– Why Logistic Regression, Classification Cost function for logistic regression, Application of logistic regression to multi-class classification, Confusion Matrix, Odd’s Ratio and ROC Curve, Advantages and Disadvantages of Logistic Regression, Case study on To classify an email as spam or not spam using logistic Regression.

**DECISION TREES AND SUPERVISED LEARNING**– Decision Tree, data set, How to build decision tree? Understanding Kart Model, Classification Rules- Over fitting Problem, Stopping Criteria And Pruning, How to find final size of Trees? Model a decision Tree, Naive Bayes, Random Forests and Support Vector, Machines, Interpretation of Model Outputs, Business Case Study for Kart Model, Business Case Study for Random Forest, and Business Case Study for SVM

**UNSUPERVISED LEARNING –**Hierarchical Clustering, k-Means algorithm for  
clustering, groupings of unlabeled data points, Principal Component Analysis (PCA), Independent components analysis(ICA), Anomaly Detection, Recommender System-collaborative filtering algorithm, Case study on Recommendation Engine for E-commerce/retail chain.

**DEEP LEARNING –**Neural Network, Understanding Neural Network Model, Understanding Tuning of Neural Network, Case study using Neural Network

**NATURAL LANGUAGE PROCESSING –**Intro to Natural Language Processing (NLP), Word Frequency Algorithms for NLP Sentiment Analysis, Case Study on Twitter data analysis using NLP.

**PYTHON FOR DATA SCIENCE AND MACHINE LEARNING**

**PYTHON PROGRAMMING BASICS** – Installing Jupyter Notebooks, Python Overview, Python 2.7 vs Python 3, Python Identifiers, Various Operators and Operators Precedence, Getting input from User, Comments and Multi line Comments.

**MAKING DECISIONS AND LOOP CONTROL –**Simple if Statement, if-else  
Statement, if-else-if Statement, Introduction to while Loops, Introduction to For Loops, Using continue and break.

**DATA TYPES : LIST, TUPLES AND DICTIONARIES –**Python Lists, Tuples, Dictionaries, Accessing Values, Basic Operations, Indexing, Slicing, and Matrices, Built-in Functions & Methods, Exercises on List, Tuples and Dictionary.

**FUNCTIONS AND MODULES**– Functions, Why Defining Functions? Calling Functions Functions with Multiple Arguments, Anonymous Functions – Lambda Using Built-In Modules, User-Defined Modules, Module Namespaces, Iterators and Generators

**FILE I/O AND EXCEPTIONAL HANDLING**– Opening and Closing Files, Open Function, File Object Attributes, Close Method , Read, Write, Seek. Exception Handling, the try-finally Clause, Raising an Exceptions, User-Defined Exceptions Regular Expression- Search and Replace, Regular Expression Modifiers, Regular Expression Patterns and Re module

**NUMPY**–  Array Creation, Printing Arrays, Basic Operations- Indexing, Slicing and Iterating Shape Manipulation – Changing shape, stacking and splitting of array Vector stacking

**PANDAS** – Importing data into Python, Pandas Data Frames, Indexing Data Frames, Basic Operations With Data frame, Renaming Columns, Subletting and Filtering a data frame.

**MATPLOTLIB –**Plot, Controlling Line Properties, Working with Multiple Figures and Histograms.