Why Big Data and Data Science

Businesses and the governments are finding ways to make sense of all the available data in this big data era. Business Analytics thus finds favor as it utilizes tools and techniques like data mining, pattern matching, data visualizations and predictive modeling to predict and optimize outcomes and derive value from the data. Equipped with this useful information, organizations can compete better in cut-throat markets both locally and globally.

Today, data is everywhere. We create it simply with the touch of a button. But how much of it is actually useful? Whether you are in finance, operations, sales & marketing or planning, you may be in touch with millions of data points every day without being aware of how to derive valuable information from this data.

Who should attend this Data Science training

This training reviews the commonly applied statistical/mathematical/artificial intelligence concepts and theories in data science applications. Discussion of the theories surrounds the fundamentals, working principles, and limitations of commonly used algorithms. In subsequent hands-on sessions, participants will learn how to apply these theories to real-life case studies. Therefore, this training is ideal for both researchers who focus on algorithms, and for professionals who intend to work on data analytics applications such as critical product analysis, targeted marketing, customer lifecycle management, social media analytics, fraud detection, and inventory management.

What you will learn on Data Science

The course will enable you to:

➢ Explore data to find new patterns and relationships (Data Mining)
➢ Predict the relationship between different variables (Predictive Modeling, Predictive Analytics)
➢ Predict the probability of default and create customer Scorecards (Logistic Regression)
➢ Understand a Problem in Business, explore and analyze the problem
➢ Use tools like R (open source) and Excel to interpret data
➢ Solve business problems using analytics (in R) in different fields

Date: 18-20 January 2016
         (Monday - Wednesday)
Time: 0900 - 1700
Venue: Level 8 (North Wing),
Knowledge Workers Development Centre
(KDC)- MDeC, Lot 12122 Persiaran Apec,
Cyberjaya 63000, Selangor
(within Tech Mahindra Campus)

HRDF Claimable

Course Outline

Day 1: Analytics, Data Science and Statistics

Data Exploration:
- Different types of analytics
- Data variables & data summarisation methods
- Relationship among data variables
- Visualisation of data distribution

Probability Concept and Statistical Inference:
- Central tendency, symmetry and skewness
- Random variables
- Formulation of probability distribution
- Central limit theorem
- Sampling and statistical inference
- Confidence intervals, Hypothesis testing

Data Downloading and Reading:
- Importing and reading data from local files, Excel files and XML (using R)
- Importing data from MySQL and the web
- Subsetting and sorting data
- Reshaping and merging data

Case Study 1: Statistical approaches to banking sector credit card data analysis using MS Excel

Day 2: Predictive modeling using Regression

Correlation and Regression:
- Analysis of Variance, Regression Analysis
- Linear regression
- Multivariate Linear Regression (Using Excel and R)
- Logistic regression (Using Excel and R)
- ANOVA, R-Square, p-values, vif, Multicollinearity
- Heteroskedasticity, Gini coefficient
- Logistic Regression: Gini coefficient, vif (Using Excel and R)

Case study 2: Production data analysis using regression (Tool: MS Excel & R)

Day 3: Data Mining with Computational Intelligence

Data Mining
Computational Intelligence:
- Fuzzy Computation, Neural Computation, Evolutionary Computation

Fuzzy Computation:
- Fuzzy sets and operations, Fuzzy logic and approximate reasoning
- Fuzzy inference models

Neural Computation:
- A neuron, Linear Separability
- Perceptron, Multilayered Perceptron
- Backpropagation algorithm
- Self-organizing map
- Fuzzy adaptive resonance theory

Evolutionary Computation:
- Genetic algorithm
- Evolution-based models
- swarm intelligence and emerging models

Improving Data and Models
- Data Transformation; Train, Validate, Test; Tuning models

Case Study 3: Improving Data and Models - application and improving of statistical models

About Go Training

Go Training applies effective pedagogical methodologies that demonstrate case studies and hands-on practical skills, in addition to explaining clearly how things work in principle. Every course that we conduct is delivered by a subject matter expert who holds the academic qualification and working experience in that specialization. On the days when they are not teaching, our trainers work on consultancy projects and technical deliveries. Their work has received numerous recognition and awards in the industry. Our team of trainers has been invited as keynote speakers at numerous international conferences, and as principal consultants for various industries.

About the Instructor

Victor Tan is working as a seasoned Data Scientist in a MNC. He is a graduate of University College London where he received a MSc in Machine Learning.

Upon graduation, Victor launched his career with a Fusionex as one of the very first hire of Data Scientists. During the period of employment, he imparted his freshly acquired knowledge hence allowing and expediting the company's transformation from a software company to a true BDA company. Being a regional BDA pioneer, the company has acquired numerous BDA solutions contracts with both the local government and private cooperation. After that, he joined a MNC where he is currently working at. In the company, he communicates and works together with Data Science experts around the world on daily basis.

Victor has experience in applying Data Science in Retail Analytics, Social Media Analytics, Image Processing, Financial Analyses, IoT and Telco CRM Analytics. He also carries out Data Engineer's and Software Developer's work. He is a CCDH with knowledge in Map-Reduce and Spark and studying for CCP:Data Scientist. He uses Matlab, R, C# Java, SQL, Python and Shell Scripts. He has won a faculty award and a hackathon competition too.