SAS - Big Data, Data Mining, and Machine Learning

This course introduces the concepts of analytical computing and various data mining concepts, including predictive modeling, deep learning, and open source integration. The course introduces a wide array of topics, including the key elements of modern computing environments, an introduction to data mining algorithms, segmentation, data mining methodology, recommendation engines, text mining, and more. Throughout the course, concepts are introduced, explained, and demonstrated using approachable real-world examples. The instructor will share his extensive experience from consulting with clients on their analytic efforts as well as from his own projects throughout his career.

Skills Gained
- use a data mining methodology
- apply modern cutting-edge algorithms to data
- implement best practices in the development and maintaining of analytical models
- explore the opportunities to create value through analytics
- assess different machine-learning models
- explain in simple terms model data mining and machine-learning methods
- incorporate unstructured data into your business models and decisions
- build segments for business applications.

Who Can Benefit
- Business analysts, statisticians, data scientists, their managers, and anyone who wants to understand what analytics is and how analytics can benefit their business

Prerequisites
- No previous courses are required. It is recommended that you have a basic working knowledge of analytics and/or statistics.
  Familiarity with the SAS language is a plus.

Course Details

The Computing Environment
- hardware
- storage (disk)
- central processing unit
- memory
Network
- distributed systems: database computing and file system computing
- virtualization

Analytical Tools
- exercise: inventory of tools

Predictive Modeling
- a methodology for building models
- SEMMA
- model types: binary classification, multilevel classification, and interval prediction
- assessment of predictive models
- common predictive modeling techniques
- RFM
- contrast predictive modeling from segmentation

Overview of Segmentation Ideas and Techniques
- segmentation
- cluster analysis
- distance measures (metrics)
- evaluating clustering
- number of clusters
- K-means algorithm
- hierarchical clustering
- profiling clusters

Common Predictive Modeling Techniques
- regression
- generalized linear models
- neural networks
- decision and regression trees

Special Applications
- incremental response modeling: building the response model and measuring the incremental response
- time series data mining
- recommendation systems
- text analytics: information retrieval and content categorization
- text mining