Advanced Methods and Applications in Data Science (42 Hours, 14 Lectures)

Description

Data Sciences is a fast-evolving practice that apply several sciences, theories and techniques in solving complex data-related problems and develop applications that support transforming the way organizations do their activities. In this course, students will leverage the concepts acquired from Introduction to Machine Learning and Statistical Analysis through hands-on applications as well as learning more advanced methods and techniques.

The course covers variety of data Science practical methods including machine learning, deep learning, data mining, statistical modeling, visualization and big data applications using proven tools like R, Python, Spark, MLlib, GraphX, Greenplum, Tablue among others. Furthermore, students will learn how the actual customer engagements in this filed works including consulting and implementation centered around generating business value out of data.

Pre-requisites

- CIT-651: Introduction to Machine Learning and Statistical Analysis

  (Or equivalent knowledge subject to NU evaluation)

Reference Textbooks

Learning From Data, Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin, March 27, 2012.

Other References including:

- Articles, Blogs, Forums, Tutorials …
- Technical Papers, Research Papers, Text Books…
- Industry Analysts, Vendors, Experts…-

Course Topics

- Big Data and Data Science Use Cases and Applications
- The Data Science Modeling Lifecycle
- The practice of the Data Sciences vs Traditional DW/BI
- Applying statistical analysis and modeling
- Visualization and Exploratory Data Analysis
- Data Integration, Quality and Implications on Big Data
- Modeling Unstructured Data and Data Streams
- Association Rules Mining
- Working with Clustering Techniques

- Linear, Logistic and Multinomial Regression
- Naïve Bayes Classifier
- Decision Trees Classifier
- Support Vector Machines
- Random Forests and Ensemble Learning
- Big Graph Theory and Network Analysis
- Time Series Analysis and Forecasting
- Deep Learning, ANN and CNN
- Data Driven Transformation for Organizations
- Deployment Considerations for the Big Data Platforms
- Consulting Skills, Agile Delivery Methods