40.016 The Analytics Edge

Time: Mondays and Thursdays (as scheduled)

Venue: the course will be delivered <u>online</u>

Course Description:

The increasing availability of data is changing the way organizations are thinking about themselves and the way they interact with the world. Smart use of data aids in improving profits in businesses, the quality of life of individuals, the performance of sports teams, and the way we network with each other. In this course, you will learn how to use analytics to give you an edge. The course will expose students to real world examples of how analytics have been and are being used. Through these examples you will learn how to use tools of predictive and prescriptive analytics such as linear regression, logistic regression, discrete choice models, classification and regression trees, random forests, clustering, and optimization. The course will also demonstrate how ethical considerations can be raised in analytics. We will use the statistical software R (for predictive analytics) and the numerical computing software Julia with optimization packages (for prescriptive analytics).

Learning Objectives

1. Identify the link between data and models to help make decisions that add value to individuals, companies, and institutions;

2. Describe data effectively, predict future outcomes, and prescribe decisions using the tools of analytics.

Measurable Outcomes

1. Develop a mathematical model from a given dataset;

2. Solve the mathematical model using the tools of analytics such as linear regression, logistic regression, discrete choice models, classification and regression trees, random forests, clustering, and optimization;

3. Justify the decisions from the mathematical model.

Textbook

- *The Analytics Edge* by Dimitris Bertsimas, Allison K. O'Hair and William R. Pulleyblank. Dynamic Ideas, Belmont, Massachusetts, 2016.

Notes and supplementary material will be provided each week. The students can also refer to the online course on edX offered by MIT - The Analytics Edge 15.071x. A related book for statistical learning is:

- *An Introduction to Statistical Learning with Applications in R* by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer, 2014.
  `https://link-springer-com.library.sutd.edu.sg:2443/book/10.1007/978-1-4614-7138-7`

Software

We will make use of the free statistical software R for the course.
Download it at `http://www.r-project.org/`.
The best way to use R is to combine it with RStudio.
Download it at `https://rstudio.com`.
We will also make use of the numerical computing software Julia for the course.
Download it at `https://julialang.org/`.

Prerequisites

Statistics and Optimization.

Course Assessment

Mid-Term Test (Week 6)	25%
Competition (Week 13)	38%
Final Test (Week 14)	25%
Participation	10%
Course Feedback Completion	2%

Instructors

Bikramjit Das. Email: bikram@sutd.edu.sg
Stefano Galelli. Email: stefano_galelli@sutd.edu.sg
Abhishek Pal Majumder. Email: abhishek_pal@sutd.edu.sg
Foo Lin Geng. Email: lingeng_foo@mymail.sutd.edu.sg
Consultation: Wednesday 2:00 to 4:00 pm, online

Topics to be covered (Tentative)

Week 1	Introduction to Analytics and the Software R with Visualization.
	Recall: Statistical tests, tools.
Week 2	Method: Supervised learning (Linear Regression)
	Predicting the quality and prices of wine (Wine analytics)
	Method: Unsupervised learning (Principal Component Analysis)
	Healthcare/Text analytics
Week 3	Method: Logistic Regression
	Predicting the failure of space shuttles (Challenger),
	Predicting the risk of coronary heart disease (Framingham Heart Study)
Week 4	Method: Multinomial Logit and Mixed Logit in Discrete Choice
	Predicting the Academy Award winners (Oscars)
	Estimating the preference for safety features in cars
Week 5	Methods: Big Data and Analytics: Model Selection
	Baseball (Sports)
	Cross-country growth regressions (Economics)
Week 6	Review and Test (23 October, Friday, 2:30 pm-4:30 pm)
Week 7	Break
Week 8	Method: Classification and Regression Trees (CART), Random Forests
	Forecasting Supreme Court Decisions (Law)
Week 9-10	Method: CART, Random Forests, Naïve Bayes Classifier
	Text Analytics: Twitter (Social media), Enron (Email)
	Ethics in Analytics
Week 11	Method: Clustering, Collaborative and Content Filtering
	Netflix, MovieLens (Recommendation systems)
Week 12	Method: Optimization
	Revenue Management, Capstone Allocation
Week 13	Review and Competition
Week 14	Test (16 December, Wednesday, 9:00 am - 11:00 am)