COURSE SYLLABUS PROPOSAL

DATA MINING WITH BUSINESS ANALYTICS

Nedret Billor
Auburn University (USA)

July 2 - July 12
(9:00 am - 12:00 m)

July 6
(9:00 am - 12:00 m)

DESCRIPTION

Massive collections of data are created by businesses, governments, and individuals as a by-product of their activity. Therefore, decision-makers and systems depend on intelligent technology to analyze data systematically to improve decision-making. In this course, our focus is on the ability to understand and translate business challenges into data mining problems and on examining how data analysis technologies can be used to improve decision-making. Therefore, we will emphasize heavily on students obtaining hands-on experience in implementing a range of commonly used data mining techniques by using "R", the widely used programming language, on business analytic problems. We will study the fundamental principles and techniques of data mining, and we will examine real-world examples and cases to place data-mining techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science.

OBJECTIVES

At the completion of this course, students will be able to:

1. Think carefully and systematically about whether and how data can improve business performance, to make better-informed decisions for management, marketing, and so on by considering business problems data-analytically

2. To interact competently on the topic of data mining for business intelligence. Know the underpinnings of data mining techniques, algorithms, and systems well enough to interact with expert data miners, consultants, Chief Technology Officers, and so on

3. Have had hands-on experience mining data

METHODOLOGY

- Topic 1.1: Preliminaries: Overview of Data Mining Process
  - Topic 1.2: Data Exploration and Dimension Reduction
- Topic 2: Prediction and Classification Methods
  - Topic 2.1: Multiple Linear Regression
  - Topic 2.2: k-Nearest Neighbors
  - Topic 2.3: The Naive Bayes Classifier
  - Topic 2.4: Classification and Regression Trees
  - Topic 2.5: Logistic Regression
  - Topic 2.6: Discriminant Analysis
  - Topic 2.7: Combining Methods: Ensembles and Uplift Modeling
- Topic 3: Cluster Analysis
- Topic 4: Forecasting Time Series
- Topic 5: Social Network Analysis
- Topic 6: Text Mining

EVALUATION

There will be three projects and one final project to be graded.

Mini Proj 1 on Regression: July 4
Mini Proj 2 on Classification: July 8
Mini Proj 3 on Clustering: July 10
Final Project: July 12

Mini Project 1: 20%
Mini Project 2: 20%
Mini Project 3: 20%
Final Project: 40%

PRE-REQUISITES

Familiar with the basic statistical concepts such as mean, standard deviation, normal distribution, t-distribution, confidence interval, hypothesis testing, p-value, etc (Reference: Neil A. Weiss (2016), Introductory Statistics, 10th ed). Previous knowledge of R is not required.

Universidad de los Andes | Vigilada Mineducación
Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964.
Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.