

ECO520: Business Analytics Tools II

Online Class

updated: 3/27/2020

INSTRUCTOR: Jin Man Lee

Office: Loop Campus (1 E Jackson DPC 6230)

Phone: (312)362-5970, Email: jlee141@depaul.edu

COMMUNICATION: The best way to reach me is to send an email. Please use ECO520 as a prefix on the subject line to get my attention. If you don't receive my reply within 24 hours, please remind me again. Due to some email filters, your email might be lost. Email is only for any personal issues. All questions related to course material and homework should be posted on D2L. You will find the right answers on the discussion in D2L.

COURSE OBJECTIVES

This course introduces advanced data analytical skills by identifying meaningful patterns of data and transforming patterns into statistical models to make more profitable decisions using big data. It covers descriptive, predictive, and prescriptive analytics. The topics are included on advanced levels of clustering analysis, discrete choice, multivariate regression, random forest, and neural network analysis to find the best suitable techniques to drive the best data analytics in business decisions. SAS and R will be extensively used to develop the best analytical models using various business data, including social media, housing, health, bank financing, and public data such as Census and CPS.

PREREQUISITE

Applied Quantitative Analysis or passed the equivalent tests at the graduate school level

RECOMMENDED TEXTBOOK

- Applied Analytics through Case Studies Using SAS and R: Implementing Predictive Models and Machine Learning Techniques, Author: Deepti Gupta, ISBN: 9781484235249, Publisher: Apress (We will use many examples from this book)
- More required readings and articles will be available on D2L every week

SOFTWARE USED in CLASS (required for Assignments)

- SAS Studio from SAS academy, SAS Studio from SAS University Edition, or SAS for Windows using apporto.com cloud service (<https://depaul.apporto.com/>)
- R for any platform. Current version: 3.6.0. This is a free public domain statistical program available to install. We will use R studio server from <http://bigblue.depaul.edu:8787>

EXAMS and PROJECT

- Midterm Exam (Take-home exam starts on May 5 to May 12 at 10:00 PM)
- Final Project Proposal (PPT) (6/2/2020 at 10:00 PM)
- Final Project Due (6/13/2020 at 10:00 PM, upload all files to D2L)

ASSIGNMENTS Most of the Assignments will be focused on computational work using data in class or data chosen by me.

- All assignments are to be prepared individually unless otherwise stated by me. You risk an academic integrity violation if you submit the same work and answers with others. Group study is encouraged but not the submission of Assignments.
- All statistical code needs to be submitted as a txt file, For example, assignment.sas.txt or assignment.R.txt.
- All submitted codes should be tested without any error. If there are any errors, the assignment will get zero credit.
- All weekly assignments should be uploaded to D2L by Tuesday at 10:00 PM
- Late submission will get a steep penalty without prior approval by the instructor.

GRADE

Weekly Assignments (45%), Midterm Exam (25%), Final Project Presentation and Submission (30%)

Scale of grade: A: 93 or above, A-: 88-92.9, B+: 85-87.9, B: 80-84.9, B-: 77-79.9, C+: 75-76.9, C: 70-74.9, C-: 68-69.9, D+: 65-67.9, D: 60-64.9, F: Below 60

ACADEMIC HONESTY

Work done for this course must adhere to the University Academic Integrity Policy. Violations include but are not limited to the following categories: cheating; plagiarism; fabrication and academic misconduct.

- Cheating: any action that violates University norms or an instructor's guidelines for the preparation and submission of assignments. Such actions may include using or providing unauthorized assistance or materials on course assignments, or possessing unauthorized materials during an examination.
- Plagiarism: the representation of others' work as your own. You are to prepare your own assignments. Violations may result in the failure of the assignment, failure of the course, and/or additional disciplinary actions.
- Misconduct: This includes but is not limited to attempts to bribe an instructor for academic advantage; persistent hostile treatment of, or any act or threat of violence against, an instructor, advisor or other students. Violations may result in additional disciplinary actions by other university officials and possible civil or criminal prosecution.

You may review the Academic Integrity Policy in the Student Handbook or by visiting Academic Integrity at DePaul University (<http://academicintegrity.depaul.edu>)

STUDENT WITH DISABILITY: Students with Disability may register The Productive Learning Strategies (PLuS) Program. You may request your exam schedule arrangement by requesting through the PLuS program. For more information on the PLuS program, you may visit <http://studentaffairs.depaul.edu/plus/> or call: 312-362-8000.

SUMMARY OF WEEKLY SCHEDULE

Here is the schedule for each week

1. Wednesday 10:00 AM : New weekly material will be posted in D2L including lecture note and homework
2. Tuesday 10:00 PM : All weekly homework Assignment Due (One time revision of homework without penalty)

TENTATIVE SCHEDULE OF TOPICS

(The instructor may change the order or contents by needs, any special material needs for the class will be available on D2L)

I. Introduction to Business Analytics and SAS

- TOPIC 1: Introduction to Business Analytics (BA)
 - Business Analytics, Data Analytics, and Big Data
- TOPIC 2: Introduction to SAS Programming
 - SAS Data procedure syntax
 - SAS Descriptive Analytics
 - Assignment 1 (SAS Coding Assignment)

II. Advanced Topics in Descriptive Analytics

- TOPIC 3: Statistics and Data Exploration for BA
We will explore the topics on more issues on descriptive statistics and data issues using SAS. Following topics will be included:
 - Statistics and Inference: Descriptive Statistics, Inference, and Hypothesis Tests on Business Decision
- TOPIC 4: Validation and Sanitation of Data
We will explore the topics on more issues on descriptive statistics and data issues using SAS. Following topics will be included:

- Data Exploration and Sanitation: Understand the challenge of real data and enhance the ability to programming skills
- Missing Value and Outlier Treatment
- Assignment 2 (SAS Descriptive Analytics and Cleaning Data)
- TOPIC 5: Supervised and Unsupervised Learning
Classification and Clustering Analysis
Clustering Analysis in Big Data: Cluster effect depending upon the geographical, political, and socio-economic environment. Since the process of identifying, defining, and describing is not standard, the student will need to work on their data to understand how to cluster their data for further data analysis.
 - K-Means
 - Hierarchical Clustering
 - Data Visualization and Statistics by Clustering
 - Assignment 3 (Clustering Analysis)

III. Predictive Analytics

- TOPIC 6: Regression Models I
Regression and forecasting techniques can yield new insight for managers by uncovering patterns and relationships that they had not previously noticed or considered. The student will learn how to make appropriate regression models reach a business decision by allowing various alternative models.
 - Simple and Multiple regression models
 - Dummy and Categorical Variables
 - Nonlinear and Functional Forms
 - Regression Diagnostic Tests
 - Prediction Power Tests
 - Assignment 4, 5 (Regression Analysis)
- Midterm Exam
- TOPIC 7: Introduction to R
 - Introduction to R for Predictive Analytics
 - Bring your own data (BYOD) proposal and data collection
 - Assignment 6 (Regression Analysis using R)
- TOPIC 8: Regression Analysis in Discrete Choice Model:
models might be used for supply chain, investment decision, response to marketing communication, payment of credit card, shoppers' brand choice model, consumer purchasing decision model, drug choice model, and readmission of patients in a hospital.
 - Linear Probability Model

- Binomial Logistic Regression
- Assignment 7 (Discrete Choice Model)

IV. Machine Learning (ML) and Deep Learning (DL) Models

- TOPIC 9: Machine Learning (ML) Models
 - Random Forest Model (RFM) is a supervised classification algorithm. We can see it from its name, which is to create a forest by some way and make it random.
 - Assignment 8 (RFM)
- TOPIC 10: Deep Learning (DL) Models
 - Artificial Neural Network Model (ANN) currently provides the best solutions to many problems in classification, clustering, and regression analysis. Image recognition, speech recognition, and natural language processing. deep neural networks as components of larger machine-learning applications involving algorithms for reinforcement learning, classification, and regression.
 - Assignment 9 (RFM, ANN)

V. Case Study and Project Management

- TOPIC 11 Project Management and Presentation
 - Case Study: Data Analytics and Strategy - Chicago Housing Price Models, Loan Performance Model
 - Descriptive, Predictive, Prescriptive Analysis using clustering, regression, and machine learning models.
- FINAL PROJECT

Each group will work on the topics of interest. This is a proposal PowerPoint includes introduction the data and descriptive analytics. The project will start after the midterm exam, and the related techniques will be developed throughout the class. If allowed, students can bring their own data or any public data they are interested in. The maximum number of members are three in a group.

FINAL PROJECT DUE : June 13 (Saturday) 10:00 PM