OPIM 5604-B13: Predictive Modeling (LEE)
Fall 2017 – Course Number 13429

Group Project Instructions

The group project is an important component of this course. You will work with your teammates on a predictive modeling project. Your group is expected to discover/predict interesting relationships and patterns using a significant amount of real data. Each group should be comprised of 4 or 5 members. Each group should choose one of the following topics:

- Predictive modeling / knowledge discovery project. You can explore/analyze/mine interesting data sets and present and describe your findings. In other words, describe: (a) the data and how it was collected, (b) what you might be looking for in such data (e.g., potentially interesting/important business questions that could be explored/answered using this data), (c) the tools/techniques that you used, (d) the results that you obtained, (e) and the conclusions that you can draw (e.g., the actions that one can take in response to the findings).
- Case study. Provide an in-depth description of a specific real-life data analytics project (e.g., undertaken by the company that you work for). What was the business problem? How was the problem addressed before the project took place? What kind of data was used and how was it collected? Which specific techniques and technologies were deployed? How did it transform the business? What were the main obstacles? What were the measures of success/failure? What are the main lessons that have been learned?

Each group has to obtain the instructor's permission regarding the appropriateness of the topic chosen by submitting a 1-page proposal by **October 11.** Your group project grade will be based on: a) white paper and b) in-class presentation.

Software

Although you are encouraged to use JMP for data analysis and prediction, you are allowed to use your favorite software package/programming language, or a combination of tools for this project.

Required Components

Whether you choose to do a predictive modeling project, case study, or a technique project, you need to have an appropriate **data set** to work with. Ideally, this data set should be large in terms of both columns (features) and rows (observations). Once you have decided on such a data set, you will follow the *SEMMA* procedure and document each step in detail in your written report. Specifically, SEMMA stands for

1. Sample (Sample from the original data; partition into training, validation, and test datasets)
2. Explore (Examine your data both visually and statistically)
3. Modify (Transform variables if needed, detect outliers, manage missing values)
4. Model (Fit different predictive models)
5. Assess (Assess model performance and compare models using a validation dataset)

This framework allows you to conduct a predictive modeling task in a systematic manner. You are required to document each SEMMA step in your report: **provide screenshots and statistical test results if applicable**. Chapter 2 of the textbook by Shmueli et al. provides a useful guide to the process of data mining and predictive modeling.

(Note: If you are working on a case study project, you will also need to obtain relevant data set and follow SEMMA for your analysis.)

Most importantly, please be aware of any **existing confidentiality or non-disclosure agreement** that might be in place—make sure you have the permission to use, analyze, and present the data.

Evaluation of Your Project

Assessment	Percentage Weight
Project Presentation	**8% (Presentation) + 2% (Peer Evaluation)**
White Paper	**8% (Write-Up) + 2% (Peer Evaluation)**

Each group should upload a group member sheet to HuskyCT on Sep 20, which lists the group members (4 or 5 people).

Each group should submit a 1-page **proposal** on Oct 11, which describes (1) the topic chosen, (2) a brief description of the project including the data source and summary statistics of the data set.

Each group will have 15 minutes to present their project in class on Wednesday, Nov 29. This 15-minute presentation should cover (1) motivation and problem definition, (2) analyses performed, (3) main findings, (4) implications and actionable plans, and (5) brief Q&A. This presentation is 10% of your course grade.

In addition to the in-class presentation, each group must submit a white paper for your project. The white paper must contain an executive summary of the problem and major findings, a list of actionable recommendations, data analytics performed (a documentation of the SEMMA process), as well as all other supporting materials. The white paper is 10% of your course grade.

Important Dates

Sep 20: Group Member Sheet due (Deliverable**: group member list uploaded to HuskyCT**)

Oct 11: Proposal due (Deliverable**: proposal uploaded to HuskyCT**)

Nov 15: Group project workout session--check in with instructor individually. (Deliverable: n/a)

Nov 29: Presentation day (Deliverable: **presentation slides**, **white paper, both uploaded to HuskyCT**)