

Course Syllabus: CSC 672 Predictive Analytics Capstone

Winter quarter 2016/17

Instructor:

Raffaella Settimi, PhD

Office: CDM 716

Email: rsettimi@depaul.edu

Tel.: 312 362-5556

The best way to contact me is through email. Most emails will be answered within 24 hours. You can also call me in my office at 312 3625556. If you want to meet with me, please either come or call during the regular office hours or email me to set up an alternate meeting time.

Description:

This capstone course provides an opportunity for students in the Predictive Analytics program to integrate and apply the analytics skills and knowledge learned in the previous courses to a large data analytics program. This is a largely self-directed course, where my role as instructor is to provide guidance and suggestions to each team.

Students work in teams and may propose their data analytics project. At the end of the course, teams submit a report summarizing their analyses and study outcomes, and present results to the class.

Grading

Since the focus of the course is to work on a data analytics project in a team, your grade will depend on the following factors:

Team grade:

Project deliverables (50%)

Presentations (10%)

Individual grade:

Paper presentation and discussion (20%)

Attendance and active participation (10%)

Weekly status reports/blogs (5%)

Performance evaluation by your peers (5%)

Class Format

There will be meetings scheduled throughout the quarter to discuss status of ongoing project. The assigned class meetings are on Monday at 5:45-9:00pm, but I expect to hold meetings at other times depending on the team members' availability. In addition to the meetings with the instructor, each team is expected to hold weekly meetings among themselves to discuss and work on their respective tasks.

I will be available during office hours to have one-on-one sessions with the students on issues related to the project. Meetings with online students can be done via video conferencing (through google hangout or skype). Each student is expected to sign up for at least one meeting each quarter. Students can also make appointments with me if they are not available during the official office hours.

Readings

Throughout the quarter students will be assigned papers to read and discuss in class or on the online forum. You will receive 10 points for reflections of the readings which you will present in class (in-class students) or post in the class forums for online students. You will receive an additional 10 points based on your response and discussion in class and/or in the forum to other posts. A critical component of the class design is collaboration among peers. If you do not actively participate in the discussions in a meaningful way, you will not get full points.

Group project

Teams

Teams should have no more than four members. You can find the list of students (both inclass and online) with email information in the Content page to facilitate the creation of the groups. Please post group information and "group name" on the discussion forum.

Each team must elect a team leader that will serve as a contact point for the group. The leader will also make sure that the deadlines are met, and deliverables are submitted through D2L.

If you need help finding a team, email me as soon as possible, and I'll identify a team you can join.

Requirements

I will set up a group site on D2L, but teams can also create their own Google group if that's more convenient.

The class forum will be used to evaluate the team and team members contributions to the project. Each team member should post one entry each week to discuss his/her contributions to the team work.

Entries should include:

- Tasks accomplished this week
- Problems or issues encountered in the analytics task
- Possible ideas for solving issues
- Work plan for next week
- Weekly status report/blogs are due on each Sunday at 11:59pm.

Project status and presentations of intermediate findings

There will be milestones throughout the quarter to evaluate your work progress.

Milestone 1: Team information and analytics project proposal

Deadline: Monday January 9th by noon

Proposal should include:

- List of team members
- Name of your team leader who will be the contact point for the team and be responsible for the timely submission of deliverables.

- If your team is selecting a project provided by me, you should tell me which project you will be working on. If your team wants to use a different dataset, send me a proposal for the project including a description of the data, data source, number of variables, size of dataset, research questions.

Milestone 2: Preliminary analysis

Deadline: Sunday January 22nd by midnight

The cleaning and preparation of the data will take a considerable amount of time, but this is an essential task in your analytics work. Aggregate different data sources and apply preprocessing techniques to prepare the data for your analytic task.

- Apply data exploration techniques (data visualization, descriptive statistics, etc...) to learn about the scope and quality of your data
- Are there missing variables or outliers? Decide what you will do with them.
- Identify role of variables and start planning what technique to use to accomplish the analytics task

The deliverables for this milestone are

- 3-4 page document describing preprocessing steps to clean the dataset. This report should document variable transformations and recoding, analysis of outliers, evaluation of data properties and decisions about missing values. Document should also outline plans for future work.
- presentation (in class or video recorded) to describe steps.

Milestone 3: Mid project status check

Deadline: February 5th by midnight

In this milestone, each team should present the intermediate results of their project. Teams should prepare a deck of slides that describes the steps of the analysis up to this point. The goal of this presentation is to describe the contributions toward the completion of the project, and to discuss the future plans to complete the analysis.

Milestone 4: Project status check:

Deadline February 19th midnight

Each team should present the additional work that was accomplished after the previous milestone. The goal of this milestone is to present contributions toward the project, and to discuss issues and challenges encountered during the work.

Milestone 5: Final report and Presentation

Monday March 6th, 2017

This is the final milestone. Each team should prepare a presentation to describe the results of the analytics project. The presentation will be in on Monday March 6th, and I am hoping that all students will be able to participate in person or via video conferencing.

Project report

Monday March 13th, 2017

Each team should also submit a 10-12 page report describing analytics task and findings. This paper should be exceptionally well written with a clearly defined problem and detailed solution, and should include an executive summary and a technical paper.

The executive summary should be no longer than one typewritten page, describing the conclusions of your data analysis to a non-technical audience. It should be intelligible to a person who does not know data mining or machine learning techniques. Suppose you are talking to your boss or to a friend who is not familiar with statistical terminology and data science methods. This can be seen as the executive summary/introduction of your report.

A *technical report* describing your analysis and the data. You can include graphs and output tables, only if you use them in your discussion. This section is intended for a statistically literate/technical audience and must be written in a clear organized fashion. For instance, you can organize the report into subsections such as

- Data preprocessing
- Exploratory analysis of the data.
- Methods (describing main research questions and analytics methods)
- Analysis of results
- Validation and testing
- Discussion and Conclusions.
- Appendix

Each group member must submit an evaluation of the work of each group member. This evaluation must be completed by each group member and must be submitted through D2L.