CISC 6080 Capstone Project in Data Analytics
Department of Computer and Information Science
Dr. Weiss, Summer Session III 2020

Instructor: Dr. Gary Weiss (http://storm.cis.fordham.edu/~gweiss)
Office: LC 6th floor (TBD)
Email: gaweiss@fordham.edu
Office Hours: By appointment

Class meeting times: The meetings times will be agreed to by the course participants and the instructor during the initial class meeting. Some of these meetings may be one-on-one with the instructor due to the nature of the capstone course. Because the capstone project is a substantial project, students are not required or expected to complete the project during a single 5-week summer session (summer session III spans both summer sessions I and II).

Course Description: The goal of this class is to sharpen students' skills in data analytics. The student will work with the instructor to identify an appropriate capstone project, and then will implement the project. The project should be substantial and involve either research in data analytics, or address a substantive real-world data analytics problem (or a combination of both). Ideally the project should utilize many of the skills that the student has acquired in the MSDA program. The student is expected to describe the project and its outcomes in a project paper. At the end of this class students should have a deeper and more comprehensive understanding of data analytics, and have experience in addressing complex data analytics problems. This course is for graduate students only.

Textbook and Required Readings: No textbook is required but online readings will be assigned, mainly in the form of conference papers and journal articles. The student is also expected to identify related work in the project area, which should be presented and discussed within the class—and described in the final project paper.

Academic Honesty: All work produced in this course should be your own unless it is specifically stated that you may work with others. Violations of this policy will be handled in accordance with university policy which can include automatic failure of the assignment and/or failure of the course.

Grading: The percentages given below are guidelines for both the student and instructor and may be changed as needed to reflect circumstances in the course. Any changes that occur are likely to be minor. For those students working closely with a faculty member other than the instructor, the other faculty member will be asked for his/her input when assessing the project.

Capstone Project Proposal 10%
Final Capstone Project Paper 60%
Final Capstone Oral Presentation 15%
Assigned Readings (Presentations, Summaries, and Discussion) 15%

Capstone Project Details:
The capstone project should be a significant analytics project that requires substantial effort and a breadth of knowledge. It may be an applied project, research project, or some combination of the two. Capstone projects in data analytics will normally be done alone, but complex projects that warrant it may be undertaken by several students with permission of the instructor. The capstone project proposal, project paper, and oral presentation are described on the following pages.
**Project Proposal**

The project proposal should be completed near the beginning of the class to allow sufficient time for completion. Ideally the student would enter the course with some concrete ideas of projects that they would like to pursue. The instructor will help identify potential projects for students who do not already have concrete project ideas. The written capstone project proposal must be typed and should be approximately 2-3 pages long. The purpose of the proposal is to make sure that your project is reasonable and enables the instructor to provide useful feedback.

The capstone project proposal should at minimum cover the following items:

- Preliminary title and student name(s)
- Abstract: Similar to the abstract that will ultimately appear in your paper. It should be one paragraph long and provide a motivation for the project, a high level description of the project, and the project’s main expected outcomes.
- Project Description: A description of the project and the main steps necessary to complete the project. Below are some of the items that will likely be covered for all capstone projects:
  - What data sets do you plan to use? If you must do significant work to get the data or convert it into the proper format, then describe the process and approximate effort required. If you plan to collect your own data (e.g., via a survey), discuss the collection process. Describe the total amount of data that will be available and justify that it is sufficient for the project.
  - What is the main data mining/data analytics task. The description does not have to be long but it must be clear and well-defined.
  - What learning tools/toolkits do you plan to use (e.g., WEKA, Python Scikit) and what data processing methods and model induction algorithms do you plan to explore?
  - How will you assess/evaluate your results and how will you determine the utility of these results? Are there other papers/results that you can compare them against, or is this a novel problem?
  - What related work exists? Try to provide a high level description of the major related areas and related papers that fall in each of these areas. Minimally the proposal should include at least 4 related papers (the final project should include more).
- Timeline: The proposal should break down your project into major and minor work items and provide an estimate of the amount of effort required for each; a tentative timeline should then be assembled.

**Capstone Project Paper**

The project and its results should be documented in a professional manner. The instructor will provide IEEE and ACM Conference Paper Templates (both use two-column formatting) and the student is expected to use one of these for the paper. Different formats can be used with prior agreement from the instructor. The main body of the paper is expected to be between 6 and 10 pages (longer papers are fine but should be discussed with the instructor). Your paper should generally adhere to the following outline, which is typical for a conference paper, but minor variations are acceptable:
• **Abstract**: summarizes the paper and the goals of the work (required)

• **Introduction**: Introduces the project and what you are trying to do. May include some background.

• **Background**: Depending on the project, you may want a separate background section, depending on how much background you want to include. For example, it may provide domain information for the domain that you are studying.

• **Experiments**: Describes the experiments and the experimental setup. Will describe the data sets, the evaluation metrics, the data mining tools used, and any other details related to the experiments.

• **Results**: Includes the experiment results (which are typically not included in the experiments section). A discussion of the results may be included, or they could be included in a separate discussion section, which follows the results. For now, we will assume no separate discussion section.

• **Related Work**: A brief description of related work, with citations to relevant papers. Research oriented papers will typically have more references, but all papers must contain at least 6 references.

• **Conclusion**: Provide your conclusion. For example, comment on the quality of your results. You may also want to include some material on future work, whether or not you intend to do such work.

*Capstone Oral Presentation:*
Each project must be presented to the class via an oral presentation. Presentations should generally take 20-30 minutes. Visual aids (e.g., graphical results) should be provided.