

STAT 8320 Applied Multivariate Data Analysis

Final Project Description

Overview

There is a semester-long course project that is meant to be a **substantial business predictive analytics effort** related to **real-world data management issues**. The students can choose either of the following two options for their course project:

1. **Application-flavor project:** Students select their **own dataset** and **identify an interesting and nontrivial graph analytics problem**. You then need to **figure out solutions to the problem and perform thorough experimental studies to testify to the effectiveness of your methods**.
2. **Implementation-flavor project:** Students select one (or several) published paper from the data analytics journals and conferences (for example IOWA's business analytics journal list, or you can search on Google scholar) published on or after 2015, implement the core algorithms and systems specified in the paper, and carry out experimental studies mentioned in the paper. The paper's code should not be available online.
3. **Research-flavor project:** In this option, students will focus on exploring and developing novel ideas in the field of multivariate data analytics. You are expected to either extend an existing research problem or propose a new one. This can involve formulating a new theoretical model, designing a new algorithm, or developing a unique application in a real-world domain. The project will require you to perform a comprehensive literature review, define your hypothesis, and provide experimental results to validate your approach.

Students can form teams of 2 to 3 people. Students are welcome to discuss their problems, ideas, and potential solutions with the instructor, and even other faculty members throughout the semester.

Milestones

1. **Group formation (0%):** find project partners and begin to discuss project problems and ideas.
2. **Project proposal (10%):** your proposal is one or two pages long and should explicitly state the following: 1. **Your project type: application-flavor**, implementation-flavor or research-flavor (If this is an implementation-flavor project, please indicate the paper you want to implement. If it is application-flavor, please describe the dataset(s) you want to analyze); 2. **The problem your project will address**; 3. **Your project goal and motivation**; 4. **The (rough) methodology and plan for your project**. Be sure to structure your plan into a set of incremental, implementable milestones and include a schedule for meeting them; 5. **The resources needed to carry out your project**; 6. **The workload distribution**.

3. **Status report (10%):** Your status report is one or two pages long and should contain enough implementation, data, and analysis to show that your project is on the right track. You should revise your original proposal to accommodate the TA's and instructor's comments and any surprising results or changes in the direction, schedule, etc. You sometimes also need to have a refined version of the problem statement. Basically, the following items are expected in your report: 1. A very clear and specific problem you want to solve (you've finalized the problem statement so far); 2. The basic goal of the project (what do you want to achieve at the end of the semester); 3. Your assumptions and methods 4. Your software/tools/data sets used in the project; 5. The detailed plan of experimental studies you want to perform (in accordance with the experimental studies mentioned in the paper); 6. Your current status and partial results; 7. Your brief plan for the remaining month.
4. **Final project presentation (20%):** During the lecture in the last week, you will have a presentation explaining your problem statement, dataset, methods, analysis, and results. The presentation should include all the sections in the final report (please see below).
5. **Final report and software/source code (60%):** the final report should extend your previous writeups into a conference-style paper with five to ten pages (single or double-column). The report should: 1. present the problem statement and summarize your contributions in the first section; 2. include a detailed description of your algorithms, analysis, and implementation in the technical section; 3. describe the evaluation methodology and significant results in the evaluation section; 4. finally, present your conclusions (in the summary section); 5. for teamwork, the report should also include a paragraph explaining, for each group member, their contributions and duties in the project. 6. Please specify a hyperlink through which we can download your source code, software, and data set for reproducing your experimental results.

Software, source code submission: Please provide your COMPLETE source code, datasets, and runnable software in one package. You may provide a link to DropBox, Github, Bitbucket, etc. Please include a README file specifying how to install and run your software and all the experiments. Students are required to use R or Python for programming. Students can use libraries or online code during implementation, but such source code won't be considered as your workload.

Stages	Due	%
1 - Group formation	Friday, January 17	0%
2 - Project proposal	Friday, February 14	10%
3 - Status report	Friday, March 28	10%
4 - Presentation	Friday, April 25 (Analytics Day)	20%
5 - Final report and software/source code	Monday, April 28	60%

Table 1: Final project timeline (dues are all at midnight).