INFO 659 FINAL PROJECT

Intro to Data Analytics project

PURPOSE

The purpose of the final project is for you to practice the entire data analytics process end-to-end in a collaborative setting on a domain and problem of your choosing.

GROUPS

This is a group assignment; form a group of 2–4 people. I will create a thread in the Week 2 forums for people who need to connect with others to form a group.

Along with your group's submission, each individual student will submit an assessment of their own and their teammates' contribution to the project. I will consider this in setting the final grades; the default will be to give all team members the same grade, but if there is clear evidence of disparate contribution, I may adjust individual members' grades.

REQUIREMENTS

Find a domain of data and problems where you can use methods and tools discussed in this class for data analytics and problem solving. Use real (not hypothetical) data and problems related to your interests. While there are various types of data you can use, make sure there are a sufficient number of features/attributes and rows/instances for meaningful analyses. Think of problems you can help solve with the data in terms of various learning styles: association learning, clustering, classification, and numeric prediction. Eventually you will follow the data analytics lifecycle/processes in your project.

If you are having difficulty finding data sets, look at Assignment 1 for some suggested repositories of data you might want to use.

PROPOSAL SUBMISSION

After you form a group, submit a description (abstract) of about 200 words about your initial project idea, including the data you plan to use and how you will obtain it. I will need to review and approve it before you get started on the project. The proposal submission is in Blackboard as an assignment.

PROGRESS UPDATE (OPTIONAL)

Create a presentation to report your group project progress. Please review Chapter 12 of the textbook and follow the structure to develop related slides. While your project may not be completely done at the time of this presentation, please do include the following aspects among others unique to your project:

- Business domain and problem statement
- Situation and project goals
- Data overview, concept of learning
- Approach / methodology
- Data preparation, visual exploration, cleaning, transformation, etc.
- Modeling, parameters, and evaluation details
- Key points, findings, and/or recommendations
- Conclusion, future work, etc.

Depending on the nature of your project, you may conduct modeling for numeric prediction, clustering, classification, and/or association rules learning. Please include plots / visualizations to help tell the story about your project.

FINAL SUBMISSION

Final project is 30% (30 points) of your final grade. Please submit the following items for your project as a group.

- **Project presentation:** Please submit your presentation slides if you have updates.
- **Raw and processed data:** Please submit major data files (or links to them) you produced as part of your initial data collection, preprocessing, transformation, and analysis.
- **Data and documentation (analytics report):** Please document your project in terms of data analytics objectives, plan, processes, results, and findings. Here are potential sections to include in your report.
- **Introduction and problem statement:** Your overall goals and objectives for the projects and your (research) questions.
- **Data sources and data preparation:** Data sources, original formats, what is in the data, relevant variables and details. The steps you have taken to prepare and/or convert data for the proposed analytics.
- **Data exploration, visualization, cleansing and transformation:** The processes in which you explore, visualize, cleanse, and transform data for the project.

- **Methodology:** How you approach the problem and analysis, why you select the models/methods, and details of related models.
- **Modeling and results:** Processes in which you apply models to the data, details in the results, results plots / visualization, etc.
- **Evaluation:** How you interpret the results, what evaluation metrics you use, what you find and learn from result evaluation.
- **Major challenges and solutions:** Discussion of major challenges, issues, and ideas in completing the tasks.
- **Conclusion and future work:** What did you learn from this project? What else can you do in the future?
- Please also include any other aspects specific to your project.