MSE 446 Project Instructions

The goal of the project is to apply machine learning to a real-world problem of your choosing. The project may be done in groups of up to four students and is worth 30% of the course grade: 5% for the proposal and 25% for the final report and Python code.

Below are some ideas for your projects. If you have other ideas, you're welcome to discuss them with the instructor and the TAs

- Collect some data (e.g., by surveying people), try several machine learning algorithms and/or parameter settings, report business and/or scientific insights.
- Take an existing machine learning algorithm and modify it. Test your solution against the original version on several datasets and report the performance improvements (in terms of accuracy, running time or both).

Project proposal (5%), due 4:00pm Monday, March 3

Please submit one pdf per group on Learn. Page limit: two pages. Please include the following:

- Names of all team members
- **Problem statement**: What problem will you solve / What hypotheses will you test? Be specific.
- **Proposed methodology**: How will you solve the problem or test the hypothesis? Describe the data and the methods you plan to use.

Your proposals will be graded on technical quality (creative and interesting choice of problem, technical correctness, feasibility, good understanding of what machine learning is and what it can do) and writing quality (well organized, clearly and concisely written).

Project report (25%), due 4:00pm Wednesday, April 2

Please submit one Python notebook file per group, with all the text, code and figures. Please include the following:

- Names of all team members
- **Problem statement**: restate the problem you are solving or hypothesis you are testing
- **Data and methods**: describe the data and methods you will use. Include illustrations!
- **Results**: describe your solution. For each result/experiment, clearly state the purpose, illustrate the results, and describe the insights/lessons learned.
- **Conclusions**: include an overall conclusion, summarizing what you learned about your specific problem and about data analytics in general. Also suggest a few directions for future work if you were to continue this project.

Your reports will be graded on technical quality (appropriate methodology, well thought-out experiments, results clearly described and illustrated, meaningful conclusions) and writing quality (well organized, clearly and concisely written, good programming style).