MUSA 650 Project

In this project, you will solve a real problem using machine learning and satellite imaging.

Individually or in groups of at most 3 students, you will come up with a problem, find a real dataset, and implement machine learning approach(es) to solve it. You are expected to make full use of all methods, techniques, and terminology you have learned in the course. You are also expected to describe the problem, suggested approach(es), your methods and results of your experiments in a clear and well-written presentation.

We strongly encourage you to think about your problem, and discuss with us to fine-tune to make sure it is interesting, but feasible to solve.

The project will be split into five parts, all due by 11:59pm on the date listed.

- Preliminary Report (10%), due April 11th
- Progress Report (10%), due April 22nd
- Final Report (40%), due May 2nd
- Presentation (30%), due May 2nd
- Discussions (10%), due May 10th

Preliminary Report (10%) - due April 11th

Your preliminary report should give an overview of what you plan to do in your project. You should outline the problem you plan to solve, the data you plan to use (along with a description of the dataset), and at least **two** different approaches you plan to take.

The preliminary report should be approximately one page in length, and should include links to the dataset(s) you plan to use.

Progress Report (10%) - due April 22nd

Your progress report should outline the progress you have made towards your project. You should discuss what approaches you have taken so far, how well they are performing, and any issues you are currently facing.

The progress report should be approximately one page in length, and should include an appendix of any performance plots you may have ready.

Final Report (40%) - due May 2nd

The final report should summarize your entire project, from problem statement to a discussion of your results. Your final report should include the following sections:

- **Introduction**: describe the relevant background information of your problem, and motivate why it is important (including articles or papers as appropriate).
- **Data**: describe the data, where it comes from, and how it was collected. You should summarize and visualize the data as well, as appropriate. For example, you may perform PCA, or visualize images.
- **Methods**: describe your solution to the problem, and the machine learning methods you used to solve it. Make sure to explain why you picked this approach, and how it was tuned. You should also discuss your performance criterion, and how they were chosen.
- **Results**: describe your results, using plots and other visualizations as appropriate. Compare and contrast any additional methods you used. You should include all necessary data and visualizations to convincingly show that your approach has solved your problem.
- **Discussion**: discuss the results of your approach. Did your machine learning method perform as expected? Why or why not? Did you solve your original problem? How might your results be practically useful or applicable? Also make sure to discuss any other approaches you tried, and why they were not successful.

Your final report should be approximately 4 pages in length, not including plots and other media. Include your **commented** code with your report in a format where it can be ran from start to finish. You should also include references used, such as your dataset, textbooks, papers, news articles, etc.

Presentation (30%) - due May 2nd

Your presentation should summarize the information in your final report in a more visual format. You should include all sections from your final report, but *only relevant information needs to be included*. Your presentation should be *concise, informative, and visually interesting*.

The audience will be your fellow classmates, who will read your presentation and discuss it on Piazza. Your presentation should be no more than 20 slides.

Discussions (10%) - due May 10th

Individually, you should post thoughtful discussions on the presentations of **3 different groups** on Piazza. Make sure to discuss how you feel they solved their problem, including the approach taken as well as the applicability of the results.

Feel free to respond or build upon the posts of your classmates, but make sure your discussions are insightful and constructive on their own.

Each discussion comment should be at least two paragraphs long. Make sure your comments are respectful and constructive; if you disagree with a result, provide thoughtful feedback.

Submissions

All reports must be submitted via Google Forms in PDF format, just as you did with your Homeworks.

Your presentation should be posted as a public note on Piazza in PDF format, so that your classmates may see it and post discussion comments.

All parts are due by 11:59pm on the date listed.