Instructions for Machine Learning Project, Fall 2022

Deadlines:
- Project proposal is due Nov 10th before class
- If you do not receive an email from me requesting you to make changes to your proposal, you have permission to start with your project on Nov 11th.
- Final project report is due Dec 12th before midnight by submitting the pdf file and any supplementary materials through Canvas.

Project Theme

Each project is individual. You are expected to **commit 20-25 hours total** to your project and demonstrate that you are able to transfer what you learned in this course to practice. In your project, you are expected to **obtain a data set, perform exploratory data analysis, define a machine learning problem, use python or related software to perform experiments, and write a report**. You are allowed to use existing toolboxes and code you find on the Internet, under condition that you acknowledge those sources. As far as the data set, it could be something you have experience with, something you are working on at the moment as part of your research, or a data set you download from Internet. If your project is related to your ongoing research with your academic advisor, you need to ensure me that you will not just copy the existing results. So, you need to explain in what aspects your course project is different from your existing results. If you download a data set from Internet, it is very likely you can also find ready python code that applies machine learning on it. This is also not acceptable. So, you will need to ensure me that you will not simply use the existing code on the existing data.

For some ideas about the data, the following is a list of pointers to interesting data sets. Please explore those links and see if there is anything that captures your imagination. You should also feel free to use Google search and find some other sources of data. You should also feel free to crawl data from Internet of from deep web. Some pointers:
- Kaggle data sets: https://www.kaggle.com/datasets
- UCI Irvine Data Repository: http://archive.ics.uci.edu/ml/, http://kdd.ics.uci.edu/
- Links from Wikipedia: https://en.wikipedia.org/wiki/List_of_datasets_for_machine_learning_research
- More links: https://www.dataquest.io/blog/free-datasets-for-projects/
- More links: http://www.kdnuggets.com/datasets/index.html
- More links: https://blog.bigml.com/list-of-public-data-sources-fit-for-machine-learning/
- More links: https://www.analyticsvidhya.com/blog/2016/11/25-websites-to-find-datasets-for-data-science-projects/

Project Proposal

Proposal should be up to **1 page in length** with 11pt font size and single spacing and submitted as pdf file on Canvas. It should contain

- Project title and student name
- Introduction Paragraph: give motivation; describe the problem; mention any related work and results
- Proposed Work (one or two paragraphs): explain the problem; explain the proposed approach
- Timeline: propose the timeline for the project (state what will be done after week 1, 2, and for the final report)
- References: provide at least 2 related references (could be a web link)

Final Report

Project deliverables should include a project report and relevant programs and result listings.

Project report should consist of: Title, Abstract, Introduction, Methodology, Results, and References sections. Programs and results listings to be delivered will depend on the specific project. The report should be 4 pages in length with 11pt font size and single spacing. It should contain:

- **Project title and student name**
- **Introduction**: give motivation; describe the problem
- **Approach**: explain the proposed approach
- **Results**: explain what was done, show results as tables or figures, discuss results
- **Conclusion**: Summarize the whole project and its outcome

Other Deliverables:

- Any code you created or used could be submitted as ipynb file(s)
- Additional figures and tables – if you really like your results but cannot fit them in the 4-page report, place them in a separate file called the appendix. Tables and figures should be clearly labeled and with captions describing what we are looking at.
- Pointers to any artifacts related to your project that you wish to share