CS/SE 4AL3:
Project Proposal

Fall 2025
Due: October 1

1 Overview

For the course project, you will gain experience with building a fully functional machine learning application. You get to come up with this project with your teammates, however, this project should be non-trivial, and have significant effort in some combination of the data collection techniques, modeling approach, and evaluation strategy. You may not directly replicate existing projects.

2 Proposal

The first milestone is to complete your project proposal. The format of the proposal will be a PDF document which includes the following items:

1. Team members **(2-3 people required)**
2. Task title and overview, including the significance and what makes it challenging
3. Task definition (type of data, classification/regression/generation, number of classes, single label or multi-label)
4. Describe the problem, it's impact on the real world, and why it is challenging.
5. Data source(s) and plan for data collection. This may include how you are going to scrape the data and follow terms-of-service, API access and handling rate limiting, using open-source data, or any other relevant details. If your data does not have labels, how do you plan to get them? If assigning labels by hand, how long does it take per instance? Include links to the data if relevant. If you download your dataset from Kaggle, you must include the Kaggle link and the original data source link from where the dataset was downloaded and posted on Kaggle. Include any meta-data available for your corpus. If you are not collecting it yourself, include details about how the data was collected, annotated, preprocessed, etc. Do not work on the datasets for which no source information is available. Please indicate whether you will use a small subset of the data or features for your project or the entire dataset, and why.
6. Expected size of the dataset (number of data points) and 3 example data points with labels. Your dataset should have at least 1k data points. Some projects will have more or less data.
7. Proposed solution: How do you plan to go about solving this problem? It is okay to not know how the machine learning models work at this point in the class, but you should be starting to get some idea based on the lectures and assigned readings. What kind of features and target labels do you have? What kind of models might you try? You may not propose simple linear regression models for this project. Are there any existing solutions to your problem? You must indicate 2-5 sources (research papers, books, machine learning challenges online) from where your solution was inspired. Put some thought into how you would approach this. How will you know if the model is good? How will you evaluate it? Also, share the libraries you intend to use for the project.
8. A **team contract signed by all team members**. This is a description of the expectations and/or roles that team members agree upon for the semester. Part of your final project grade will be based on whether your teammates agree that you followed this contract. See next section for details.

3 Data Suggestions

There are many places to look for interesting datasets.

1. Canada Open Data for issues of public interest
2. UCI Machine Learning Repository for more classic machine learning datasets
3. Data published as part of academic papers (e.g. ICML, NeurIPS, AAAI for general ML papers, CVPR and ICCV for vision papers, ACL, NAACL, EMNLP for language papers)
4. Well documented Kaggle datasets – beware of the variation in quality here
5. Academic Torrents
6. Any website or social media data that allows scraping through an API or as specified in their robots.txt file

4 Team Contract

The Team Contract[1] specifies the overall purpose of the team, the responsibilities, and the ground rules or norms that team members agree to follow. Your contract should include:

1. What is our team's purpose or mission?
2. What are the duties/roles of each team member? What is expected of each team member?
3. How will the team handle the leadership/facilitation/management activities?
4. Anything else you think will be helpful to set the groups expectations.

The team should describe what is expected from each team member and each member should sign the contract **(print name, sign, and date)**. Your contract should not say "Person A will do all the work". It should focus more on how you will work together rather than who will do each individual task. You won't know all your individual tasks yet. When discussing how to handle conflict, you should write something about how you will meet, discuss, and resolve the conflict. You should not write "Person A will be shunned and banned from group activities". You will need to work together for this project.

5 Deliverables

BEFORE SUBMITTING THE ASSIGNMENT! You must go to one of the tutorial sections (does not matter which one) and get approval from the TA for your proposal. You need to bring them your full proposal (it's okay if it doesn't have the team contract yet). The reason for this is to prevent you from proposing a project that will be difficult or impossible to complete in one semester, or that would be trivially easy, and for you to be able to get some feedback from your TAs. This way you will not have to make any major changes between the proposal due date and milestone two. Approval does not mean you will get an A on the assignment. It simply means that the TA is looking out for any major pitfalls in your project.

After getting approval, you should submit your written proposal and team contract as separate PDFs on A2L. Zip these two files (zip, tar, gz are OK). Please add your group number to the top of your assignment. Your PDF file for the proposal should contain all of the above details in Times New Roman font size 12 with 0.8" margins and 1.5 line spacing. Your proposal should be 2-3 pages. Your team contract length is up to you. The contract serves primarily to hold yourself and your team member(s) accountable and to resolve possible future disputes.

[1] Borrowed ideas from `https://crlte.engin.umich.edu/wp-content/uploads/sites/5/2020/05/Team-Contract.pdf` and `https://uwaterloo.ca/centre-for-teaching-excellence/catalogs/tip-sheets/making-group-contracts` and you may also look at `https://cns.utexas.edu/sites/default/files/uploads/documents/2023-03/Examplegroupcontract.pdf` for ideas.

6 Requirements

You will be assessed based on the completion of the following requirements. Your proposal should contain all the required items with sufficient detail.

1. (25%) Data collection sources and plan (item 5, 6)
2. (25%) Proposed solution (item 7)
3. (15%) Team contract (item 8)
4. (25%) Team and task details (items 1, 2, 3, 4)
5. (10%) Document is well-organized and professionally presented

7 Next Steps

The next project step will be released at a later date. In the meantime, if you plan to collect your own data, you may want to start that process now. Data collection (e.g. scraping) can sometimes be time consuming and it is best to start earlier if possible.