CSC 4240 - Introduction to Artificial Intelligence
Final Project [100 points]

Throughout this course you have learned basics and methodology of developing an intelligent agent. One of the mostly widely used type of intelligent agents nowadays is the learning agent. A learning agent, as we learned in class, improves its performance after making observations about the world. A process known as machine learning. It also learns based on probabilistic models and ranges from simple to complex models. By the end of this course, you should have enough knowledge to apply Artificial Intelligent concepts and algorithms to real-world problems. The final project is intended to put you into this experience.

Project Groups

The final project will be group-based. A project team will consist of maximum 5 undergraduate students, or 3 graduate students. Please find teammates and sign up your team in this spreadsheet. You can use the group discussion on iLearn to find teammates with shared interests. The instructor may reduce or merge teams if needed. Your group should look for an idea/problem where you can use Machine Learning to study, analyze, and solve (fully or partially). It does not have to be an advanced idea, but it has to be something that you can do on your own and don't copy from an online source nor a project you already completed before.

Project Topics

You can choose a topic or task from any category you like. Areas like finance, service prediction, live sciences, audio, etc, are all accepted. You may replicate an existing idea but in a different scenario (may be an idea published in a research paper). You may also work on a simple version of an advanced idea, or just one stage of a multi-stage project. These links (Example1, Example2, Example3) provide some ideas/examples for projects. Many of them can be advanced and not fit, in-scale, for this project, but can inspire your imagination about what you can do.

Well developed ideas can be polished and published in a conference after the semester ends. They can also be a part of a bigger project the students can complete and publish about. You can at least prepare a poster and present it in a local or national venue for student projects.

Data

Data pre-processing/transformation and feature extraction is a very important step for Machine Learning practitioners in order to produce quality work. Please make sure to do some data exploration and analysis to make sense of the data you use. That includes, but is not limited to, inspecting and visualizing the data, dealing with missing data, and feature selection. You are encouraged to use Orange, it will save you time (this is a small introductory tutorial on how to use Orange for machine learning). You can also use snapshots from Orange analysis widgets in the final report. You may use the following data repositories to find the data you need:

- Kaggle
- UCI Machine Learning Repository
- Google Research Datasets

Methodology

The goal of this project is to test your ability to use the AI concepts we learned in class to develop Machine Learning models using the methods we studied in class and any other methods you maybe familiar

with from other courses or projects. There will be a lecture dedicated for best practices to develop machine learning models. You will **NOT** need to implement your algorithms from scratch, rather use existing libraries/frameworks of your choice. For the scope of this course and purpose of this project using Orange is acceptable, however you are welcome to use libraries like Scikit-Learn and TensorFlow. You are expected to do some self-learning to familiarize yourself more about the machine learning algorithms you are going to use. That will help you adjust/tune your model's hyper parameters to reach best performance. You will also need to provide reasonable justifications for your design and technical choices like what data pre-processing you applied, and how you chose your model hyper-parameters. A bonus is possible for exceptional work.

Evaluation

Evaluation of your project will be based primarily on the final project report and presentation. Evaluation criteria are:

- **Technical quality**: This includes using the right methodology and providing reasonable justifications for the steps taken. This also includes the problem-solution-fitness, that's whether the presented solution answers the questions presented in the chosen problem.
- **Sufficient Scope**: Does the project have sufficient scope and work for the team size. Graduate students are expected to present a bit more advanced work.
- **Originality**: Does the final project avoid being a mirror image of existing papers/projects with no net new work.

Mentoring

You are encouraged to stop by the instructor's and TAs' Office hours for guidance and advice on your design choices. You are also welcome to schedule a meeting with the instructor to discuss progress attained so far.

Grade Distribution

The final project grade will be distributed as follows:

1. Report + Code (if any): 50%
2. Presentation: 25%
3. Peer evaluation: 25%

Project Deliverables and Timeline

1. Project Teams Formed **(Due Nove 8)**
   Unassigned students and small teams will be merged
2. Topic Selection **(Due Nov 13)**
   Each team should prepare a one-page project summary that includes the following information:
   - Group participants
   - The topic/task you plan to address. Please provide a reference to the source of the idea if applicable.
   - Source of the data that will be used
   - Machine Learning algorithms you intend to use. At least 2 to compare and why. Simple and advanced ones are welcome.
   - Role of each team member. Each student should take part in the actual work of the project.

The instructor will review your proposal and provide feedback.

3. Project Report + Code (if any) **(Due Nov 27, 9am)**
   Each team should prepare a **maximum** 3-page project report (will take points off for extra pages). Your report should include:

   - The full development pipeline: data collection/generation, data transformation/pre-processing, feature extraction/engineering, data split, model tuning, etc.
   - Performance evaluation (numbers and illustrating figures)
   - A detailed section about the contribution of each team member.

   If you used Orange, then you should include the ows file (Orange Work Sapce) in the submission. If you had to write code, then you will need to submit a zipped folder with the project code. Project code should be originally yours. I understand that you will search and copy how to use certain functions and features, but anything else should be yours. Copied code will be penalized. If the data size is not huge then please include it, otherwise you may just include a link to where you got it from in the report. In both cases, you need to include a README file to instruct us on how to run your code (plus any needed dependencies) or any needed information to use your ows file. Project submitted with no instructions will be penalized.

   Any major changes to the proposed project summary (submitted by Nov 13) should be reported to the instructor for approval as soon as possible.

4. Final Presentation **(In class, weeks of Nov 27)**
   Your final presentation should summarize your work to your class-mates. The instructor(s) and other students will contribute to the evaluation of the final presentations. Date and schedule of the final presentations will be announced later.