

**SOFE4620U**  
**Data Mining and Machine Learning**  
**Project Information – Winter 2023**

- **Objective:** Apply machine learning algorithms to a real-life application. The group project is an important part of this course since it enables you to gain experience and apply concepts and technologies and utilizes many of the concepts presented in class.
- **Teams:** each team should be 3-4 students.
  - All teams (regardless of size) are expected to produce a project of equivalent scope, so you should only work in a team of cooperating people.
  - Each group should have a team leader who is responsible for managing the work of the group.
  - Any member who is failing to deliver what is required from him/her on time should be given a warning and if there is no response from him/her should be excluded from the group. You need to contact the TA or the professor to remove that member.
  - If a group decided to keep a non-participating member and puts his/her name on the report, all members will be penalized heavily (may get a zero in the project) and will be considered academic misconduct.
- **Deliverables:**
  - Project report per phase (one for the team).
  - Project presentation and final report (one for the team, with contributions from all members).
  - Every report should include a contribution matrix listing all group members and all project tasks and the percentage of contribution of each member.
- **Choosing a topic:**
  - It is a good idea to identify a topic by choosing an application domain that interests you. Then, explore how to apply machine learning algorithms that you learned and acquired from reviewing the literature to solve it. Be brave and feel free to propose ambitious things that you're excited about. If you are not certain what would make a good project, we encourage you to contact the TA by email or during his office hour.
  - Let the problem drive your choice of machine learning technique, rather than the other way around.
  - Alternatively, challenge yourself by choosing a problem or set of problems and then develop a new learning algorithm (or novel variant of an existing learning algorithm) to solve it.
  - Regardless, most projects will combine aspects of both applications and algorithms. Your project must include an evaluation on real-world data (i.e., not a toy domain or synthetic data).
  - Many fantastic course projects will come from students choosing either an application that they're interested in, or picking a field of machine learning that they want to explore more.
  - Alternatively, pick an existing open source ML project, investigate it so you can improve it by using the tools covered in class. The target will be to produce new contributions in code, technique, analysis, application, or datasets that are of value to others, while also being of value for your own education.
  - **Major topic fields:**
    - Computer Vision
    - Natural language processing.
    - Audio classification
    - Athletics and Sensing Devices
    - Audio and Music
    - Finance and Commerce
    - General Machine Learning
    - Life Sciences
    - Physical Sciences
    - Theory and Reinforcement Learning
- **Milestones:**
  - **Project proposal: [10%], due on Feb. 6<sup>th</sup>, 2023.**

- It is a 1-2 pages single spaced and 12-point font that include: Project title, names of all teammates, and a description of what you plan to do.
- You should write a compelling proposal that describes your project in detail and demonstrates your understanding and ability to complete it.
- Discuss the sources of real-world data that you will use.
- You may need to read ahead and look at other projects and ideas.
- Include citations of the articles that you used in your proposal.
- Your proposal should demonstrate that you've started to think through the various issues involved with your project and present a compelling argument in support of it.
- **Final Presentation: [30%], due on March 22-24, 2023.**
  - Prepare a 5-minute presentation video and submit it with the slides on Canvas.
  - Present your work in front of the class.
  - You must be present to get the mark of the presentation.
- **Final Report: [60%], due on March 22, 2023.**
  - Your final project report can be at most 10 pages long (include all text, appendices, figures, references, and a link to your GitHub page).
  - At a minimum your final report must describe the problem/application and motivation, survey of related work, discuss your approach, and describe your results/conclusions/impact of your project and future work.
  - It should include enough detail such that someone else can reproduce your approach and results.
  - Your report must also include a figure that graphically depicts the major component of your project (e.g., your approach and how it relates to the application, etc.).
  - A detailed README document on how to install and execute your project.
- **Evaluation Criteria:**
  - Clarity of the report, design, and results.
  - Bugs found and fixed; possible future enhancements
  - Organization of reports, including clarity of introduction, logic of structure and navigability
  - Quality of code, including coding style: comments, modularity, etc.
  - Technical quality:
    - Does the technical material make sense?
    - Are the proposed algorithms or applications clever and interesting?
    - Appropriateness of solution to the stated problem
    - Do the authors convey novel insight about the problem and/or algorithms?
    - Did the authors choose an interesting or real problem to work on, or only a small toy problem?
    - Is this work likely to be useful and/or have impact?
    - What is the level of functionality, creativity in developing the system, and its usability?
  - Novelty of the work: Is the proposed application and approach novel?
  - Clarity of presentation: Is the presentation clear? Could we reconstruct the method entirely from the report?

[Rasa.Khosrowshahli@ontariotechu.ca](mailto:Rasa.Khosrowshahli@ontariotechu.ca)

[Rasa.khosrowshahli@ontariotechu.net](mailto:Rasa.khosrowshahli@ontariotechu.net)