CSCI 374 — Machine Learning and Data Mining
Oberlin College — Fall 2022
<u>Group Project Assignment</u>

Important Dates

Assigned: October 3

Group Formation Deadline: Friday October 14 (11:59 PM)

Proposal Due: Friday October 28 (11:59 PM)

Check-in Due: Wednesday November 30 (11:59 PM)

Final Presentations: Tentatively December 5-9 (during class)

Final Implementation Due: Monday December 12 (11:59 PM)

Final Paper Due: Monday December 19 (4:00:00 PM) [when the final exam would have ended]

Assignment

In this assignment, students will collaborate in groups to practice the application of machine learning and data mining in a semester long project. For some groups, this might entail applying the algorithms and approaches studied in class to real-world problems. For other groups, it might entail studying and applying additional methods or topics within machine learning and data mining not covered in class lectures and homework.

As part of their project, each group of students will be required to:

1. **Pick a group name,**
2. **Choose a project** related to the students' personal interests,
3. **Write a proposal** identifying the problem of interest along with a proposed solution,
4. **Develop a solution** solving the problem of interest,
5. **Write a final report** describing their accomplishments, results analysis and future work, and
6. **Present the final outcomes** of their project.

The goal of this project is to provide students with an opportunity for in-depth practice applying machine learning and data mining to realistic problems related to the students' interests.

Group Formation and Choosing a Topic

The project will begin with students forming groups of 3-4 members all willing to work together on a mutually agreed upon topic. Students are strongly encouraged to discuss with one another their interests (e.g., after class, by Blackboard email) to form their groups. While forming a group, the group should decide on a project that interests each of its members.

Once a group is formed, please email the professor (aeck@oberlin.edu) with both (1) a list of your group members, (2) a group name, and (3) a 2 sentence summary of the project you have in mind (i.e., listing the topic and potential problem). This email must be received by Friday October 14 at 11:59 PM.

If you have problems finding a group or coming up with a topic idea, please let the professor know early so that we can work together to find a group and start your project.

Project Proposal

Once the group has been formed and a topic selected, the next step is to draft a formal 3-5 page (single spaced, single column) project proposal outlining:

1. A **discussion of the problem** motivating the project
   [What do we want to address with this project?]

2. A **short literature review** (3-5 sources with appropriate citations) addressing how this problem fits in with the relevant literature
   [How have others looked at this problem, if someone previously has?]

3. A **description of the data** involved in the problem and addressed in the solution
   [What data will we use in our problem?]

4. A **proposed solution** intended to solve the identified problem
   [How do we want to solve the problem involving the above data?]

5. An **outline and description of the project components** necessary to complete the project
   [What steps do we need to accomplish to solve our problem?]

6. A **description of any existing progress** towards completing the project
   [What have we done so far?]

7. A remaining **timeline** for the successful completion of the project
   [What do we still need to do, and when will we do it?]

Importantly, each proposal should contain enough details to motivate the project and demonstrate that it can be feasibly completed before the final deadline. However, given that it is only a proposal, it will probably contain some uncertainty and lack all of the complete details of a final report.

Students are ***strongly encouraged*** to work with the professor (especially during office hours, by email, or by meeting at other scheduled times if necessary) to define their projects and draft their proposals to find a project suitable for this assignment. The project should be substantially larger than a homework assignment since it is group-based and semester long, but it should also not be so complex that it cannot be finished in time.

Helpful tip: keep in mind that much of the proposal can be reused in the final report (e.g., sections 1-5 described above), so the better your initial proposal, the less work you will have to do later in the semester.

The project proposal is due Friday October 28 at 11:59 PM by email to the professor (aeck@oberlin.edu).

Final Presentation and Report

At the end of the semester, students will report on the final outcomes of their projects. This includes both a class presentation (tentatively during the final full week of class before the reading period: December 5-9) and a written report (due during finals week).

The final report should be a significant extension of the project proposal with the final solution fully explained (including its various components) with enough detail that the project could be replicated by an interested reader. Groups should also discuss:

- The work taken to create the solution and complete the project, including any **challenges or issues faced** and the **steps taken to overcome those challenges**
- The **implications of their project** (e.g., conclusions inferred from data analysis, possible practical uses of the solution), and
- **Additional future work** that would continue the project.

The format of the final report should be as if students were submitting their work to a scientific journal. The particular format is dependent on the practices of an appropriate journal venue selected by the groups.

Grading

Each group project will be graded as follows:

10% Project Proposal
15% Project Check-in
15% Final Presentation
30% Final Report
30% Solution Implementation

Between Fall Break and the end of the semester (on Wednesday November 30), each group will also turn in a "Check-in" progress report that reflects on the progress made so far and a discussion of what still needs to be done.

Each student in a collaborative group will receive the same grade on all parts, unless one or more students fail to participate compared to the more active group members. If there are problems with group participation, group members are encouraged to contact the professor during the semester so that any issues can be resolved to the benefit of all students.

Final Exam Replacement

Please note that the final report replaces the final exam for this course. The report is due when the final exam would have ended (Monday December 19 at 4:00 PM), per the Final Exam Schedule from the Registrar's Office.

Honor Code

Projects are expected to be the original idea and work of the group's members, done during the semester for the explicit purpose of this class (e.g., not a substantial reuse of assignments from other courses, private readings, or Honors projects). However, if there is a possible overlap with other work, please discuss with the professor to find an appropriate balance between the different purposes of the work. If students have an interest in machine learning or data mining that they are exploring outside of this course, it might be possible to leverage that work for this project.

Given that this is a group project, students are not only encouraged but required to collaborate and work together to complete the project. Each student is expected to contribute equally to the total work required to complete the project, and each student is also expected to contribute at least some non-zero amount to each component of the project (i.e., the design of the project, the proposal writing, the final presentation, the solution implementation, and the final report). For example, the work should not be divided such that one student creates the entire implementation, while another writes the entire final report. Likewise, presentations should involve all group members. One of the goals of this project is to provide students opportunities to work on multiple skills that will be useful in their careers (e.g., teamwork, presenting ideas, summarizing work, analyzing results, planning future improvements). At the same time, groups should feel free to divide labor so that the strengths of the different members are appropriately used. For example, the solution implementation and report writing do not need to be split exactly equally between all members.

Depending on the project, and with the professor's written permission, groups might be able to make use of existing machine learning and data mining libraries and software (e.g., Weka, Rattle, scikit-learn, Tensorflow, PyTorch) within their projects. Please meet with the professor early in the semester if this is of interest to your group.