# CS412 Machine Learning, Spring 2024: Term Project

May 13, 2024

## 1 Problem Description

The project aims to increase your knowledge about machine learning and get hands-on practical experience. The course project will allow you to work on a real machine learning problem. You will work in groups of 4-5 people that you already formed for HW4.

The task is to assign a severity score to a given bug description. Each bug report includes a description and its bug type. Bug severity should be scored from 1 to 6, from Enhancement (1), Minor (2), Normal (3), Major (4), Blocker (5), and Critical (6), in order. Below is a snippet of the data. The models will be evaluated based on macro precision.

The sniplet of the data is as follows. Download the data under the Assignment section on SuCourse.

| bug id | summary | severity |
|--------|---------|----------|
| 365569 | Remove workaround from bug 297227 | normal |
| 365578 | Print Preview crashes on any URL in gtk2 builds | critical |
| 365582 | Lines are not showing in table | major |

## 2 Kaggle Instructions

1. Access the Kaggle Page:

   - Click the competition link to go directly to the Kaggle page:

     https://www.kaggle.com/t/5b56e08e22f14999ac11b0896540bfa8

2. Join the Competition:

   - Accept the terms and rules of the competition and click *"Join Competition."*

3. Set a Team Name:

   - To set up a team name, one member of the group must change their team name, which should match the group name you formed previously for homework.

   **Note:** Until teams are merged, every participant is considered a "team" by Kaggle even if they are individually registered. Therefore only one member is required to change their team name. That member will form the team on Kaggle by inviting others to their team.

4. Invite Other Members:

- The group member (leader) who changed their team name with the group name should send invitations to other members.

- In order to send invitations, the other group members must have already joined the competition.

5. Accept the Invitation and Start Competing:

- Once an invitation is received, you can see it under the "Team" section.

- Join the leader's group and start competing.

**Important:** Joining teams is a one-time action, so make sure to join the right team. You should form teams immediately after joining the competition.

**Daily Submission Limit:**

- Each day, there's a set number of submissions allowed, so use them carefully.

- This limit applies to the entire team. The daily submission limit is three and if one team member submits once, the entire team can only submit two more times that day.

You can visually access information on these steps in the slides we have provided. You can also refer to the Kaggle documentation: https://www.kaggle.com/docs/competitions.

# 3 Project Report and Code Submission Instructions

After the competition ends, you will have extra days to submit the report.

You will submit a Jupiter notebook for your code. You should also include a PDF report. This should describe the steps you have taken in a technical report format. It should have the following format:

- Introduction: A quick summary of the problem, methods, and results.

- Problem description: The description of the problem. Formally define how you cast the problem, classification, regression, etc.

- Methods: Description of methods, you may use subsections.

- Results and Discussion: The results of applying the methods to the data set. Include the questions your experiments are designed to answer, the details of the experiments, and your observations. For example, you will be evaluated on AP, including the precision-recall curves. For example, report your findings if you tried different feature processing techniques. Interpretation and discussion of the results.

- Appendix: A clear description of the contribution of each person. You may also include extra material if you have many plots and tables.

# 4 Grading

- 50 % calculated based on the leaderboard.

- 30 % the approach you have taken. Everyone will get full credit but points will be deducted for mistakes, faulty experimental evaluation set ups etc.

- 20 % Report/code quality. Proper organization of the report, clarity, and attention to detail in describing the method and the results. Use proper tables and figures to communicate your results.