Cairo University - Faculty of Engineering
Computer Engineering Department
CMP4040 – Machine Intelligence
Spring 2024

CMP4040 Project Document

Project Description:

The goal of this project is to apply machine learning to a real-world problem. First, you will have to find a problem that can be formulated as a machine learning task and find a dataset for it. Then you will have to apply at least three machine learning algorithms on the dataset aiming to solve the problem.

Phase 1: Project Proposal

In this phase, you should select a problem and a dataset. The problem should satisfy the following conditions:

1. It should not be a trivial problem. For example, you can look for problems that are presented in contests or that are being researched in recent peer-reviewed papers.
2. It should have a publicly available dataset.
3. It should not be a problem that you worked (are currently working on) on in any other course or project. Consequently, it should not be part of your graduation project.

Then you should write a proposal which contains the following:

1. The team number and a list of member names (alongside their section and bench numbers).
2. The selected problem. You should clearly state the problem definition and motivation.
3. The evaluation metrics.
4. Links to the dataset and any references (e.g., links to papers, contests, etc.).

If you are unsure whether the problem you selected is appropriate or not, you can propose up to 3 problems (ordered from your most to least preferred) and we will select the first acceptable problem.

Phase Deliverables: The proposal document (PDF).

Phase Deadline: Saturday, March 23rd, 2024, 23:59.

Phase 2: Project Implementation and Report

In this phase, you should apply machine learning with the aim of solving your selected problem. Try to document your steps from the start as it will help you in writing the report. The workflow should be as follows:

1. **Analyze your dataset**. The goal is to familiarize yourself with the dataset before diving into the upcoming steps. Try to visualize the dataset, build histograms of features or outputs, look at

random samples and look for outliers, etc. It is also useful to use a baseline (e.g., ZeroR) to put your results into context.

2. **Apply at least 3 different machine learning methods** on your selected problem. The methods should be selected from the ones covered in the course. None of the methods are allowed to be deep learning. For each method, you should test the effect of the hyperparameters and try to find a set of hyperparameter values that work well for your problem.
3. You should **apply all the concepts you learned in the course** such as Generalization and Combating Overfitting.
4. Report your findings in a clear and concise manner.

Phase Deliverables:

1- Project Report (PDF) which should contain:
   a. The team number and a list of member names (alongside their section and bench numbers).
   b. The contribution of each team member.
   c. The problem definition, motivation, and evaluation metrics, including links to the dataset and any references. (This part can be copied from the proposal).
   d. Your results (the dataset analysis results and the experimental results).
   e. A discussion of your experimental results (an in-depth analysis).
   f. Your conclusion.

2- The project source code.

Phase Deadline: Saturday, May 4th, 2024, 23:59.

Rules:

1- Each team can consist of up to 4 members.
2- Any evidence of plagiarism will result in receiving ZERO points for the project.
3- Each member in the teams will be graded individually for their contribution and understanding.