Machine Learning - Final Project
due at noon on December 9, 2020
along with the presentation and source code

Aylin Caliskan*
*George Washington University, aylin@.gwu.edu

Abstract—**A summary of the project, its motivation, the ML question, evaluation and results.**

Keywords—*Machine Learning; keyword2; keyword3*

I. INTRODUCTION

Use the provided LaTeX template and style files: https://www.overleaf.com/read/rrpfptvrtkyx The final project report should not exceed four pages (excluding references). **If you do not follow the formatting guidelines, 10 points will be deducted.** The project is about both implementing and using. If you can make a case that your work is a substantially novel scientific contribution, it is OK to not implement the algorithms. Otherwise, you should be implementing the algorithms you want to use. Clearly state if you have implemented the algorithms or used libraries and justify the choice.

Final Project Evaluation. You have one month to work on your final project. Your final project presentation, the report, and source code will be evaluated based on the following criteria. Your presentation should explicitly include all of the following information/sections in this template. Submit your code, report, and presentation by noon on December 9, 2020. Each group will present their project during class. The length of the presentation will depend on how many groups we will have. Each presentation will be followed by questions.

Sign up for a presentation slot with your group and ML project information at https://docs.google.com/spreadsheets/d/1mUMe6a4SkKYHFPgPVvmacENN1-TVn5x6Jp0G89FL2mE/edit?usp=sharing. You can update this information in the following four weeks. **Your group should have at least four students and at most seven students.**

10 points	Section I: Clearly state the machine learning problem
10 points	Section II: Related work - Brief literature survey
10 points	Section III: Data collection and/or dataset/corpus
10 points	Section III: Dataset visualization (table of statistics, scatterplot, etc.)
10 points	Section III: Data pre-processing
20 points	Section IV: Approach (Is this a novel application, are you implementing a novel method? Include a table, chart, or diagram to present the results.)
20 points	Section V: Evaluation of how successful your method or approach is. This depends on your problem but one example is having ground truth data for evaluating accuracy in supervised learning. Include a table, chart, or diagram to present the evaluation.)

10 points Section VI: What is the take away? Discuss the findings. Did you gain any insights about the problem you investigated? What are the next steps?

A. Problem Statement

Please clearly state your machine learning problem or task and explain the motivation behind trying to solve this problem.

II. RELATED WORK

Has this problem been solved before? What is different in your approach and why do you think it is the right time to focus on this problem? Please go over related work in a few paragraphs by comparing your results and findings to other approaches. You should properly cite the related work and methods. Enter latex bib entries to final_project.bib and use the \cite{bib entry keyword} command. One example is citing my bias in AI paper [1].

III. DATASET

Give details about the dataset. Why did you choose this dataset? Why is this dataset the most appropriate one for solving the problems you are interested in¿ What are the properties of the dataset? Provide a visualization of dataset properties such as statistics in a table or a scatterplot of features in 2D space, etc.

IV. APPROACH AND EXPERIMENTS

Include details about which machine learning approaches you chose and justify why this is the correct approach for the problem you are solving.

Which experiments did you run and why? What are these experiments evaluating?

V. EVALUATION

How do you evaluate your approach? What are the evaluation metrics and why are these appropriate metrics for this problem? What are the results from the experiments?

A. Subsection

VI. DISCUSSION

Discuss the results. What do these results mean and what are their implications? If you are not getting positive results, what might be causing this? What is future work in this direction?

VII. Conclusion

A brief conclusion about the findings from the problem you explored with your approach and experiments.

References

[1] A. Caliskan, J. J. Bryson, and A. Narayanan, "Semantics derived automatically from language corpora contain human-like biases," *Science*, vol. 356, no. 6334, pp. 183–186, 2017.