

Machine Learning I - (ANLY 530): Final Project

Jonathan Wayne Korn @ HU

August 17, 2020

Final Project Instructions

The purpose of the project is to learn how to formulate a problem statement or research question, determine how to best find a solution to the stated problem or answer to the research question, and develop a final written report and presentation. The project is solo or team-based, your choice. Max 4 individuals in a group. Individual grades will include points for how well they contributed to the team effort.

The course project has two deliverables:

- (1) Presentation (*10 – 15 slides*)
- (2) Written Paper (*max 10 pages*)

Each of these deliverables will be described in the paragraphs below:

The objective of this project is to apply some machine learning algorithms to address a classification or regression type problem. Develop a problem statement/hypothesis and locate a dataset that will enable you address the problem. Make sure to split the data into training and testing sets in-order to train the machine learning algorithms.

The written report and presentation should cover virtually everything about the machine learning project. It should cover the situation, problem or challenge that requires attention, the relevant background, related work, data, and technical details of the analysis, conclusions and possible directions for future work. It is recognized that not all of the following sections will pertain to each report. However, it is strongly recommended that these section topics be used as a guideline for your final project presentation and report. Presentations can follow your written report in text and graphical content.

Presentations are due @ the 3rd Executive Session, if applicable, otherwise as indicated on the course canvas page. The written report is due @ the date indicated on the course canvas page.

The two deliverables should address the following:

- (1) Introduction, motivation and general description of the situation, problem or challenge.
 - What is the situation, problem or challenge you are addressing?
 - What preliminary examination leads you to believe analytics could help?
 - What are the shortcomings of the current work/analysis that analytics could help with?
- (2) Related work.
 - Provide a thorough background for the project; e.g. about the situation, problem or challenge, about other companies that have undergone similar situations, problems or challenges and how they handled them or did not, etc.

- How does this project relate to other work that has been done on this situation, problem or challenge?

(3) Data

- Give a complete description of the data you use during the project, including any you reject.
- Provide a detailed description of your data.
- Provide any exploratory data analyses you complete.

(4) Technical Approach

- Give a detailed description of the process for your entire project.
- Given a detailed description of your approach to the algorithm you have proposed. You do not have to describe well known approaches themselves, e.g. linear regression. You do have to describe how you applied the approach you used.

(5) Test and evaluation

- Describe how you test your approach to ensure that it is valid.
- Discuss the validity of your approach.
- Describe how you will evaluate your results and/or conclusions including any specific metrics, output data, completed analyses, etc.
- Discuss the baseline you will use to compare your results to.
- Discuss how well your approach worked to address the situation or challenge, solve the problem or answer the research question.
- Discuss any potential future work. For example, if you were not able to resolve the situation or problem or answer the research question what will it take to do so? What else needs to be done?
- Evaluate and report whether or not someone unfamiliar with your work could accurately replicate it.

The written report and presentation style will be graded using the rubric below. Make sure to use APA or IEEE format. Must include at least 5 references in the written paper.

Table 1: Rubric

Deliverable	Score 100 to 90	Score 89 to 80	Score 79 to Below
Appropriate Use of Techniques	Techniques are used appropriately and result in discovery of new information, patterns, etc. as appropriate.	Techniques are haphazardly applied and only achieve basic/fundamental data analysis.	Techniques are only sometimes used correctly and do not achieve accurate and/or reliable results.
Problem Statement and Organization	The problem statement or research question is clearly organized, accurately stated and well supported.	The problem statement or research question has sufficient organization including a reasonably clear thesis and supporting details.	The problem statement or research question is rambling and unfocused. Major topics and supporting arguments are presented in a disorganized and unrelated way.
Critical thinking/problem solving	Concepts, assumptions, inferences, and conclusions are clearly and thoroughly expressed. Analysis is logical and thorough.	Concepts, assumptions, inferences, and conclusions are expressed clearly in most cases but are not expressed thoroughly. Analysis is mostly logical but may be absent or flawed in some places.	Concepts, assumptions, inferences, and conclusions are unclear, may be absent or flawed logic may be present. Analysis is minimal or absent or the logic used in argument may not be discerned.
Use of Course Material	All relevant course material is used. Specific information from readings is incorporated into analysis and critique.	Some of the relevant course material is used. General information from readings and lectures is incorporated into analysis and critique.	Minimal course material is used. Information from readings and lectures is not incorporated into analysis and critique.
Written Communication	Writing is excellent. Word usage, spelling, grammar, and punctuation are excellent.	Writing is sufficient. Adequate use of wording, grammar, and punctuation. Errors are minimal.	Writing is poor. Significant deficiencies in word use, grammar, punctuation, and/or presentation.
References/citation system	All sources are cited correctly and thoroughly (in text and on reference page); An accepted citation system is used consistently and correctly.	All sources are cited, the majority cited correctly (in text and on reference page); An accepted citation system is used correctly for majority of citations.	Some sources are cited correctly (in text and on reference page); An accepted format is not used or used for a minority of citations. Or, no reference page is present.