

STAT 380 Final Project

STAT 380 Project Final Report Handout and Rubric

The final project report is your final deliverable for this project and for your course. Congratulations on getting here! Your report is structured somewhat like a research paper, and describes what you have done and why, what your results are, and an interpretation of your results. The document has a word limit of maximum of 2000 words. However, a concisely-written document is preferred. The word limit is hard: There is a 1% penalty for every word in excess of the 2000 limit. Please count the words in your document, compute the penalty, and put it on the front page. These are not included in the word count, nor are figures or references. While figures and references are not included in the word count, please do not game the system by calling a chunk of text a “figure”. There will be a penalty if you do so.

How to submit

Mar 29: Submit topic idea

Apr 12: Draft EDA. I.e, the Illustration / Figure, Background & Related Work and Data Processing part. Upload PDF on Canvas

May 4: Upload a single PDF file on Canvas. Only one member of the group needs to submit.

Rubric

Introduction (2 points)

A brief description of the motivations behind your project, the goal of your project, why it is interesting or important, and why machine learning is a reasonable approach.

(Note that the expectations are higher here than in the project proposal, since you had more time to revise.)

- 2/2 An introduction that clearly describes the project goal, why the project is interesting and/or useful, and convincingly describes why machine learning is an appropriate tool for the task.
- 1/2 The introduction describes the project, but is vague or has information that is factually incorrect.
- 0/2 The introduction does not make it clear what the specific goal of your project is.

Illustration / Figure (2 points)

A figure or a diagram that illustrates the overall model or idea of your project. The idea is to make your report more accessible, especially to readers who are starting by skimming your work. For the project, taking a picture of a hand-drawn diagram is fine, as long as it's legible. PowerPoint is another option. You will not be penalized for hand-drawn illustrations – you are graded on the design and illustrative power

- 2/2 A well thought-out figure that communicates the core idea of your project and architecture immediately.
- 1/2 An illustration that does the job, but is not particularly clear, or possibly too wordy.
- 0/2 The illustration is significantly lacking in some respect, or contain factual inconsistencies or inaccuracies.

Background & Related Work (2 points)

A description of 1-2 related work in the field, to provide reader a sense of what has already been done in this area, e.g. papers or existing products/software that do a related thing.

- 2/2 Briefly describes 1-2 prior work related to your project to put your project into context. Your descriptions need not be complete, but should contain important work
- 1/2 Background that has omissions or factual incorrectness, but otherwise places your project into context.
- 0/1 Background contains too much information not related to your project, or has major omissions of content provided to you by your instructor or TA., or does not sufficiently put your project into context.

Data Processing (4 points)

Describe the data that you have collected and cleaned. Be clear and specific when describing what you've done, so that a classmate can reproduce your work. Show some statistics and examples of your data.

- 4/4 Clearly describes sources of data, and the steps you took to clean and format your data. Statistics and data example are well-chosen, and gives readers a “feel” for your data.
- 3/4 Mostly clear description, but some aspects of the data processing steps are vague. Statistics and data example are somewhat illustrative/helpful.
- 2/4 Vague description or missing key information about where your data comes from or what you did. No example data shown, or the ones shown are not illustrative.
- 1/4 Incomplete information.

Architecture (4 points)

A description of the final model. Do not describe all the intermediate models that you have tried. Instead, present the model (or models) whose quantitative results you will show. These should be your most interesting models. Be as specific as you can while being concise. Readers should be able to reproduce a model similar enough to yours and obtain a similar performance.

- 4/4 Clear and concise description of your model architecture, so that a classmate can reproduce a model similar to yours that will perform similarly.
- 3/4 Good description of your model architecture, but with either not enough detail to be reproducible, or too much unnecessary detail not useful for reproducing your model.
- 2/4 Some issues with the description (inconsistencies, factual inaccuracies)
- 0/4 Unclear description of the type(s) of neural network model that you will use, or a choice that is inconsistent with your problem.

Baseline Model (2 points)

Describe a simple, baseline model that you will compare your neural network against. This can be a simple model that you build.

- 2/2 A reasonable choice of baseline, accompanied by a description of the baseline so that a knowledgeable classmate can find, reproduce, or build a similar version.
- 1/2 An adequate description of a reasonable baseline.
- 0/2 Poor choice of baseline inconsistent with the problem.

Quantitative Results (4 points)

A description of the quantitative measures of your result. What measurements can you use to illustrate how your model performs?

- 4/4 Insightful, well-chosen measurements that illustrate how your model performs.

- 3/4 Minor issue with the choice of measurements, or the way the result is presented.
- 2/4 Major issue with the choice of measurements, or misleading presentation of the results.
- 0/4 No result presented.

Qualitative Results (2 points)

Include some sample outputs of your model, to help your readers better understand what your model can do. The qualitative results should also put your quantitative results into context (e.g. Why did your model perform well? Is there a type of input that the model does not do well on?)

- 2/2 Insightful, well-chosen outputs that illustrate how your model performs. It is clear how you determined which outputs to show, and why.
- 1/2 Some issues with the choice of outputs, or the way the result is presented.
- 0/4 No result presented.

Discussion (4 points)

Discuss your results. Do you think your model is performing well? Why or why not? What is unusual, surprising, or interesting about your results? What did you learn?

- 4/4 Insightful interpretation of the results that is specific to your project. Exceeds expectations.
- 3/4 Sound interpretation of the results.
- 2/4 Some issues with the interpretation.
- 0/4 Discussion does not interpret results, only repeats it.

Ethical Considerations (2 points)

Description of a use of the system that could give rise to ethical issues. Are there limitations of your model? Your training data?

(Note that the expectations are higher here than in the project proposal.)

- 2/2 Thoughtful consideration of ethical issues discussed in class, applied to your model.
- 1/2 Some consideration of ethical issues in data collection, but missing key elements.

Project Difficulty / Quality (4 points)

A measure of how “difficult” the project is, and how well your model performs given the difficulty of your problem. If your problem is more difficult than what one might expect, you should clearly articulate why in the body of your report.

- 4/4 Creates a model that is that performs better than expected, and/or is more complex than expected. demonstrates learning beyond the requirements of (e.g.) the labs.
- 3/4 Meets the expectations of the difficulty of the project, and the performance looks adequate. A poor model performance is justified.
- 2/4 Project is “too simple”, or does not perform as well as expected.
- 1/4 Below expectations. There are a variety of ways to increase your project complexity, even after the fact. For example:
 - Try different models
 - Try averaging predictions for those different models
 - Use data augmentation techniques, and discuss results

Structure, Grammar & Mechanics (8 points)

We are looking for a document that is easy to follow, grammatically correct, and well-written.

References should be in the IEEE documentation style, described here: <https://iee-dataport.org/sites/default/files/analysis/27/IEEE%20Citation%20Guidelines.pdf>

- 8/8 Clear, concise and well-written document. Exceeds expectations.
- 7/8 Well-written document that could be more concise or less error-prone.
- 6/8 Well-written document that has some issues with grammar, mechanics, or structure. Meets expectations.
- 5/8 Reasonably-written document with grammar, mechanics, or structural issues.
- 4/8 Document has many issues.