
Insert title here

[A link to your repo must be right here!!](#)

Group members

- Pelé
- Diego Maradona
- Johan Cruyff
- Roberto Carlos
- Franz Beckenbaur

Abstract

This section should be short and clearly stated. It should be a single paragraph <200 words. It should summarize:

- what your goal/problem is
- what the data used represents
- the solution/what you did
- major results you came up with (mention how results are measured)

NB: this final project form is much more report-like than the proposal and the checkpoint. Think in terms of writing a paper with bits of code in the middle to make the plots/tables

Background

Fill in the background and discuss the kind of prior work that has gone on in this research area here. **Use inline citation** to specify which references support which statements. You can do that through HTML footnotes (demonstrated here). I used to recommend Markdown footnotes (google is your friend) because they are simpler but recently I have had some problems with them working for me whereas HTML ones always work so far. So use the method that works for you, but do use inline citations.

Here is an example of inline citation. After government genocide in the 20th century, real birds were replaced with surveillance drones designed to look just like birds^[1]. Use a minimum of 3 to 5 citations, but we prefer more ^[2]. You need enough citations to fully explain and back up important facts.

Remember you are trying to explain why someone would want to answer your question or why your hypothesis is in the form that you've stated.

Problem Statement

Clearly describe the problem that you are solving. Avoid ambiguous words. The problem described should be well defined and should have at least one ML-relevant potential solution. Additionally, describe the problem thoroughly such that it is clear that the problem is quantifiable (the problem can be expressed in mathematical or logical terms), measurable (the problem can be measured by some metric and clearly observed), and replicable (the problem can be reproduced and occurs more than once).

Data

Detail how/where you obtained the data and cleaned it (if necessary)

If the data cleaning process is very long (e.g., elaborate text processing) consider describing it briefly here in text, and moving the actual cleaning process to another notebook in your repo (include a link here!). The idea behind this approach: this is a report, and if you blow up the flow of the report to include a lot of code it makes it hard to read.

Please give the following information for each dataset you are using

- link/reference to obtain it
- description of the size of the dataset (# of variables, # of observations)
- what an observation consists of
- what some critical variables are, how they are represented
- any special handling, transformations, cleaning, etc you have done should be demonstrated here!

Proposed Solution

In this section, clearly describe a solution to the problem. The solution should be applicable to the project domain and appropriate for the dataset(s) or input(s) given. Provide enough detail (e.g., algorithmic description and/or theoretical properties) to convince us that your solution is applicable. Make sure to describe how the solution will be tested.

If you know details already, describe how (e.g., library used, function calls) you plan to implement the solution in a way that is reproducible.

If it is appropriate to the problem statement, describe a benchmark model^[3] against which your solution will be compared.

Evaluation Metrics

Propose at least one evaluation metric that can be used to quantify the performance of both the benchmark model and the solution model. The evaluation metric(s) you propose should be appropriate given the context of the data, the problem statement, and the intended solution. Describe how the evaluation metric(s) are derived and provide an example of their mathematical representations (if applicable). Complex evaluation metrics should be clearly defined and quantifiable (can be expressed in mathematical or logical terms).

Results

You may have done tons of work on this. Not all of it belongs here.

Reports should have a **narrative**. Once you've looked through all your results over the quarter, decide on one main point and 2-4 secondary points you want us to understand. Include the detailed code and analysis results of those points only; you should spend more time/code/plots on your main point than the others.

If you went down any blind alleys that you later decided to not pursue, please don't abuse the TAs time by throwing in 81 lines of code and 4 plots related to something you actually abandoned. Consider deleting things that are not important to your narrative. If its slightly relevant to the narrative or you just want us to know you tried something, you could keep it in by summarizing the result in this report in a sentence or two, moving the actual analysis to another file in your repo, and providing us a link to that file.

Subsection 1

You will likely have different subsections as you go through your report. For instance you might start with an analysis of the dataset/problem and from there you might be able to draw out the kinds of algorithms that are / aren't appropriate to tackle the solution. Or something else completely if this isn't the way your project works.

Subsection 2

Another likely section is if you are doing any feature selection through cross-validation or hand-design/validation of features/transformations of the data

Subsection 3

Probably you need to describe the base model and demonstrate its performance. Probably you should include a learning curve to demonstrate how much better the model gets as you increase the number of trials

Subsection 4

Perhaps some exploration of the model selection (hyper-parameters) or algorithm selection task. Generally reinforcement learning tasks may require a huge amount of training, so extensive grid search is unlikely to be possible. However exploring a few reasonable hyper-parameters may still be possible. Validation curves, plots showing the variability of performance across folds of the cross-validation, etc. If you're doing one, the outcome of the null hypothesis test or parsimony principle check to show how you are selecting the best model.

Subsection 5

Maybe you do model selection again, but using a different kind of metric than before? Or you compare a completely different approach/algorithm to the problem? Whatever, this stuff is just serving suggestions.

Discussion

Interpreting the result

OK, you've given us quite a bit of tech information above, now it's time to tell us what to pay attention to in all that. Think clearly about your results, decide on one main point and 2-4 secondary points you want us to understand. Highlight HOW your results support those points. You probably want 2-5 sentences per point.

Limitations

Are there any problems with the work? For instance would more data change the nature of the problem? Would it be good to explore more hyperparameters than you had time for?

Future work

Looking at the limitations and/or the toughest parts of the problem and/or the situations where the algorithm(s) did the worst... is there something you'd like to try to make these better.

Ethics & Privacy

If your project has obvious potential concerns with ethics or data privacy discuss that here. Almost every ML project put into production can have ethical implications if you use your imagination. Use your imagination.

Even if you can't come up with an obvious ethical concern that should be addressed, you should know that a large number of ML projects that go into production have unintended consequences and ethical problems once in production. How will your team address these issues?

Consider a tool to help you address the potential issues such as <https://deon.drivendata.org>

Conclusion

Reiterate your main point and in just a few sentences tell us how your results support it. Mention how this work would fit in the background/context of other work in this field if you can. Suggest directions for future work if you want to.

✓ Footnotes

1. [^]: Lorenz, T. (9 Dec 2021) Birds Aren't Real, or Are They? Inside a Gen Z Conspiracy Theory. *The New York Times*. <https://www.nytimes.com/2021/12/09/technology/birds-arent-real-gen-z-misinformation.html>

2. [^]: Also refs should be important to the background, not some randomly chosen vaguely related stuff. Include a web link if possible in refs as above.

3. [^]: Perhaps the current state of the art solution such as you see on [Papers with code](#). Or maybe not SOTA, but rather a standard textbook/Kaggle solution to this kind of problem

Start coding or generate with AI.