PROJECT

Machine Learning Capstone Project
A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW
CODE REVIEW
NOTES

SHARE YOUR ACCOMPLISHMENT! 🐦 📘

Requires Changes

8 SPECIFICATIONS REQUIRE CHANGES

Dear student

Really nice work on this project. This is easily one of the better reports that I've seen in terms of content and ingenuity. I've asked you to update a few things, but I think you'll see that they are all extremely minor and mostly cosmetic. Additionally, I've made a few other suggestions throughout in case you are thinking of publishing your results later on. Almost there...keep going!

Cheers!

Definition

Student provides a high-level overview of the project in layman's terms. Background information such as the problem domain, the project origin, and related data sets or input data is given.

Excellent overview of your problem. I think your description is very clear and compelling.

Some minor suggestions:

- It's a good idea to cite some of the papers where the techniques that you're using were pioneered. This gives credit to the inventors and it shows that you know the field.
- It's also recommended that you cite some studies where machine learning has been applied to similar problems. This gives you a basis for arguing that your approach is innovative or robust (by comparison).

The problem which needs to be solved is clearly defined. A strategy for solving the problem, including discussion of the expected solution, has been made.

I like that you broke this section up into separate 'problem' and 'solution' parts. I think that walking the reader through the logic behind how you've structure the problem is potentially very useful.

Suggested:

- This is a good point to begin to justify why your solution is a good 'fit' for the problem. If you were submitting this to a journal for peer review, you'd want to keep the readers focused on what you want them to think about. If they get distracted, they can ask for random things in subsequent revisions (which can significantly drag out the process and lead to arguments).

Metrics used to measure performance of a model or result are clearly defined. Metrics are justified based on the characteristics of the problem.

The use of accuracy is well-justified.

Analysis

If a dataset is present, features and calculated statistics relevant to the problem have been reported and discussed, along with a sampling of the data. In lieu of a dataset, a thorough description of the input space or input data has been made. Abnormalities or characteristics about the data or input that need to be addressed have been identified.

Excellent job explaining the structure of the dataset and noting how to interpret the raw data. You've also taken the opportunity to begin using the data exploration to justify how you'll process and handle the data later (which is excellent!).

Still required:

- Please be sure to note the total number of samples in your data.

A visualization has been provided that summarizes or extracts a relevant characteristic or feature about the dataset or input data with thorough discussion. Visual cues are clearly defined.

Really nice job here...just one small thing. Please be sure to put labels on all of the axes. Also, figures where you use color (green for in, red for out) should have a label that indicates what the colors mean.

Strongly suggested:

- I would recommend adding figure titles to each figure (Figure 1, Figure 2 etc.). This will make it much easier to refer to your data in the text ('See Fig1(a)...'). Journals will require this and it will make your report look even more polished.

Also suggested:

- I'd also recommend making the text on the figures a bit larger so that the labels are easier to read.

Algorithms and techniques used in the project are thoroughly discussed and properly justified based on the characteristics of the problem.

Very nice!

> A value much grater

Small typo here. No big deal:)

Student clearly defines a benchmark result or threshold for comparing performances of solutions obtained.

> As a benchmark model I will use the results of our current implemented solution which is not using any probabilistic model and uses simple heuristics to detect the event.

I think this is perfectly fine, but can you explain in a few additional sentences how this model works (exactly)? It should be clear to the reader what the benchmark represents, so that they can interpret any differences between the benchmark and your solution.

Methodology

All preprocessing steps have been clearly documented. Abnormalities or characteristics about the data or input that needed to be addressed have been corrected. If no data preprocessing is necessary, it has been clearly justified.

Fair enough!

The process for which metrics, algorithms, and techniques were implemented with the given datasets or input data has been thoroughly documented. Complications that occurred during the coding process are discussed.

Very nice job here!

Please be sure to:

- Include a brief mention of any difficulties that you encountered during the coding process.
- Make sure that all the figures in this section have labels (blue dots and green dots).

The process of improving upon the algorithms and techniques used is clearly documented. Both the initial and final solutions are reported, along with intermediate solutions, if necessary.

Still required:

Please be sure to report the accuracy for the models (before and after refinement).

Results

The final model's qualities — such as parameters — are evaluated in detail. Some type of analysis is used to validate the robustness of the model's solution.

The final results are compared to the benchmark result or threshold with some type of statistical analysis. Justification is made as to whether the final model and solution is significant enough to have adequately solved the problem.

> This means that, for our samples, given we try to collect 50% of the signals for the IN label and the other 50% with the OUT label, the accuracy of the previous version of the software is about 50%.

Was this the benchmark model? Please be sure to clarify this a bit. It should be clear what the benchmark performance was so it's clear that you've beaten the benchmark.

Conclusion

A visualization has been provided that emphasizes an important quality about the project with thorough discussion. Visual cues are clearly defined.

Very nice visualization here! Please be sure to add a label explaining what red and green represent (both background and dots).

Suggestion:

- When you publish your research, it's a good idea to keep in mind that there may be some readers who are red-green colorblind. When possible, it is recommended that you use blue/yellow rather than red/green color schemes so that more readers will be able to understand the figures.

Student adequately summarizes the end-to-end problem solution and discusses one or two particular aspects of the project they found interesting or difficult.

> Another aspect of this project was that I had no out of the box ML algorithm that could be reasonably applied to the problem. I had to invent a solution and I wanted to start with a simple one that could be implemented in a short amount of time and that I could understand well. My thought was: "if this works then we are done, otherwise this will be our new benchmark model". The data is not easy to separate into classes. This ruled out logistic regression and SVMs. Having only one feature, classi cation trees are not useful. Besides, all these algorithms can't take advantage of the time series property of this data. Neural Networks may work but using them with this sort of data was clearly more complex than my working hypothesis. I'm going to describe more about this possibility in the next section. The idea of using a probabilistic model was in my opinion the one best tting the problem.

I would recommend putting this earlier in the report (maybe the solution statement section). It's excellent critical thinking and it really justifies the approach that you took here. Also, this is your chance to brag a little bit about the unique aspects of your solution. Don't feel shy!

Discussion is made as to how one aspect of the implementation could be improved. Potential solutions resulting from these improvements are considered and compared/contrasted to the current solution.

Great job! One thing that might help with give you some better results in the tradeoff between spurious classifications and delayed classifications would be to write another algorithm that learns to classify based on the momentum of M. This is a common strategy in investing and I think it would pay off really well here, given that your scale on the y-axis is exponential (unless you need to worry about someone slowly creeping backwards and forwards as they enter a room).

Quality

Project report follows a well-organized structure and would be readily understood by its intended audience. Each section is written in a clear, concise and specific manner. Few grammatical and spelling mistakes are present. All resources used to complete the project are cited and referenced.

The report is very well-written (one of the best that I've read in many ways) and displays great critical thinking throughout.

> I proofreaded only the capstone PDF report. Please forgive me for any typos you may encounter in the ipython notebook.
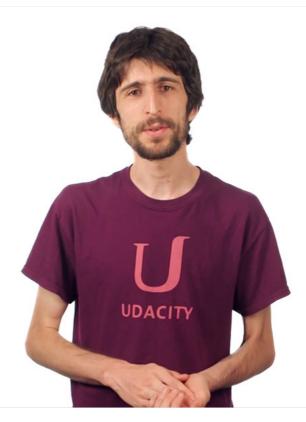
Your English is excellent!

Code is formatted neatly with comments that effectively explain complex implementations. Output produces similar results and solutions as to those discussed in the project.

Your notebook is quite excellent! However, please be sure to add some comments to the evaluate_series() and build_model() functions. These are the complex implementations that you've coded from scratch. A new reader who isn't super familiar with the field could probably use some help understanding what you're doing.

☑ RESUBMIT

⬇ DOWNLOAD PROJECT

Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

⊙ Watch Video (3:01)

RETURN TO PATH

Rate this review