UDACITY

Machine Learning Capstone Project

REVIEW
CODE REVIEW
HISTORY

Requires Changes

4 SPECIFICATIONS REQUIRE CHANGES

Overall fantastic work! There are just a few sections that need a little more information, but you have an incredible foundation here and the project was sincerely great to read.
Keep up the great work - you're nearly there!

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.

Cool idea! The problem space is presented clearly and some background information has been provided. Awesome work here 👍

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.

GREAT

Great start here - you've definitely shown that that this is a problem suitable for ML.

You'll want to discuss your solution proposal in a little more depth. Think about things like:

- What sort of models are you planning to use?
- How do you plan to train these models?

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

Great metric choice here!

- F1 is definitely a reasonable candidate for this problem.
- You've justified the selection sufficiently.

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.

- The target feature was discussed thoroughly and clearly.
- Inputs are well explained, both in what they represent and in their type.
- Great job discussing the abnormalities in the data - this gives us a very thorough understanding of what the data is like.

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.

Awesome choice of visualization! This really adds to our understanding of the data.

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

Awesome job in this section - it was a genuine pleasure to read.

- Thorough description of the algorithms that you're using.
- Relevant sources are given.
- Each algorithm is reasonably justified.

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

Your benchmark is clear and sensible for this problem.

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.

Fantastic work here. Not only are your preprocessing steps well-described, but they're described thoroughly enough that an informed reader could replicate your work. This is exactly what we're looking for here - well done.

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.

You've got a great start in this section, but a little more information is necessary.

- How did you implement these models? What technologies did you use?
- What hyperparameters were selected initially?

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.

You should show the results of your grid search here - this will help a reader understand the intermediate steps between your initial and final model. A visualization will be very helpful.

As a tip, I would highly recommend also tuning the *underlying classifier* in AdaBoost. You can do this by using the `base_estimator__` in your grid search like so:

```
param_grid = {"base_estimator__criterion" : ["gini", "entropy"],
              "base_estimator__max_depth" :  [1, 2, 3, 4, 5],
              "n_estimators": [1, 2]
             }
```

In my experience, tuning the base estimator can lead to very high performance gains.
In general, it's not really necessary to tune `n_estimators` - you'll find that the more estimators you use the

better your final result will be (as a trend with some variation). There are diminishing returns on this trend, so typically you'll want to keep increasing it until the performance gain is tiny or non-existent.

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.

Overall awesome job here - this is nearly perfect.

Just make sure you talk about the other hyperparameters of AdaBoost too - although `n_estimators` is the only real hyperparameter for the AdaBoost as an ensemble, the hyperparameters used in the underlying decision tree also need to be discussed. What were the hyperparameters of these weak learners within AdaBoost?

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.

Nice work! You clearly had useful model as an end result. Congratulations on your success!

Conclusion

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

Very good choice of a visualization! It's no surprise that the number of backers is predictive of success - one might argue that a high number of backers is even synonymous with success. I wonder how well a model might work without that information?

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

Great discussion of the end-to-end process here - it's clear that a lot was learned.

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.

Very thorough discussion of possible improvements. All of these are great choices - optimizing XGBoost in particular may give you an even stronger model.

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.

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

☑ 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)

Student FAQ