UDACITY

⧉ DISCUSS ON STUDENT HUB ›

Machine Learning Capstone Project

REVIEW

CODE REVIEW

HISTORY

Requires Changes

1 SPECIFICATION REQUIRES CHANGES

Really awesome submission!
Seriously, this was very well-written and the visualizations are all beautifully presented. There's just one section that needs some more detail, but you clearly have a strong grasp of the machine learning methodology. I doubt the revision will take you long.
Keep up the great work! We look forward to the next one!

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.

Great start! This is a solid introduction to the domain, and it's clearly a problem suited to ML. The only thing I'd recommend including is some existing research as it pertains to targeted marketing - this is a very common use case for machine learning, so there is likely existing solutions that you could pull from (or draw inspiration from).

✓

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

ROC definitely makes sense for this problem. Good justification and nice job explaining how it works.

✓

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.

Very solid problem statement - your ultimate goal is clear and you've provided a solid solution strategy.

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 of the data or input that need to be addressed have been identified.

Really excellent work!

- The features are discussed in their meaning and representation
- Abnormalities in the data are thoroughly discussed
- Input space is provided
- Summary statistics are provided

🔁

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

This section needs just a bit more depth. Ultimately, the "algorithms and techniques" section is where you describe not just what algorithms you're using but also what these algorithms are and how they work. To that end, you'll need to go into more depth in describing how logistic regression and boosting work. You should cover these algorithms thoroughly:

- What is the training process for that algorithm?
- What is the prediction process?
- What hyperparameters exist? What do they mean?

Feel free to include diagrams or mathematical equations here if they help your explanation.

✓

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.

Excellent choice of visualization!

✓

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

When in doubt, a simple logistic regression model is often a very good benchmark.

Methodology

✓

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.

✓

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.

Again, excellent work. Your steps are all extremely clear, and definitely could be replicated by a competent researcher.

The elbow method is fine, but I prefer more mathematical justifications for determining the optimal number of clusters. You may want to check out silhouette scores or bayesian information criterion - both are great tools for providing some mathematical, objective measure of how effective a clustering algorithm is.

✓

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

Excellent discussion of your preprocessing steps. You've gone into really awesome detail providing your step-by-step process here. I have no doubt that someone could replicate your preprocessing pipeline based on this section.

You might be interested in checking out featuretools - this is a cool framework for automatically performing some feature engineering. It can take a lot of the legwork out of that phase of the development cycle in some cases.

Results

✓

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.
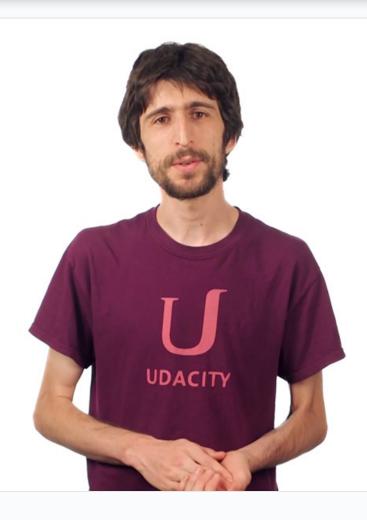
✓

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.

Really awesome analysis here! This is a really cool visualization as well - this is a format I might start using to illustrate final model results in my own reports.

☑ **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

Best practices for your project resubmission

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

▶ Watch Video (3:01)