Toggle navigation

UDACITY

- Logout
- Logout

SHARE YOUR ACCOMPLISHMENT

Project

Machine Learning Capstone Project

A part of the Machine Learning Engineer Nanodegree Program

- 4
- IONS 1
- NOTES

SHARE YOUR ACCOMPLISHMENT!

Requires Changes

9 specifications require changes

Good job! I see a lot of work in this project.

I really like the first half of the report, which is very formal and professional. But the rest of the report seems to lack of information and have some mistakes. Please see the comments and annotations for detail.

Nice trying this time. Remember that all the hard work will pay back. Keep working on it, and you are one step closer to the Nanodegree! :)

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.

Interesting project. It would be better if you can also mention a little bit information about the data, and explicitly describe what you're trying to do.

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 nice description of the project. I really like the way you divide the problem into several steps and sections. This is a very good strategy. Well done!

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

It's a little bit tricky turning a multi-classification problem into a binary one. Although detailed explanation is provided about how the metrics work, a simple figure with an example would help people better understand them.

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.

Nice data description and exploration. Are they any abnormalities or missing values in the dataset? If not, the data must have very good quality. In reality, we always have all kinds of abnormalities or missing values in our datasets :)

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.

I'm a little surprised about the number of "0" bird in the dataset.

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

Because the input is very import for the machine learning algorithm, could you give more details (maybe with an example, or some formulas) about how the single fixed-length feature vector is generated?

Here, you should also mention some information about the machine learning algorithms you're going to use. For example, what algorithms will be used? Why are they chosen? What are the cons and pros?

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

Could you provide more information for this score? For example, how is it generated? And what algorithms are used to produce this score?

Typically, we only need one benchmark for a project.

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.

Again, are there any abnormalities in the dataset?

The process for which metrics, algorithms, and techniques were implemented wi
input data has been thoroughly documented. Complications that occurred during
discussed.

Rate this review
( ) ( ) ( ) ( ) ( )

Could you give more information about the algorithms you used (e.g., pros and cons
the parameter settings for the algorithms?

In the beginning, you mentioned this is a classification problem, but here you used some regression algorithms. Could you explain the reason?

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.

A detailed documentation about how the experiments are implemented is required in order to meet specifications. A parameter table with corresponding model performance would be great. Here, you also need to compare the model performance between initial and final solutions.

Here is a potential method you can try.

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.

Similar as above problem, choice of parameters should be justified. Same for the robustness of the model's solution.

A possible way of doing this is using certain sensitivity test, and find out whether the model results are sensitive to certain parameter or not.

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.

Some type of statistical analysis is required to verify the robustness of the final result. It's a little bit tricky for the Kaggle competition.

Conclusion

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

I understand some plots have been provided. Could you think of another plot here, or move one plot to this section in order to meet specification?

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

This is a challenging project. Good job so far!

Rate this review
( ) ( ) ( ) ( ) ( )

Discussion is made as to how one aspect of the implementation could be improve[illegible] resulting from these improvements are considered and compared/contrasted to the current solution.

XGBoost, which is a common used algorithm in Kaggle, could be another useful method for this project.

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 first half part of the report is really nice. However, the last half part starts to have some problems, such as typos and lack of information. Note that, large code snippets should be avoided in the formal report. They can be attached as separated files. Details can be found in annotations.

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

Comments are very important for coding. They can help you remind your thoughts for part of the code after a long time, while also help other people understand your codes.

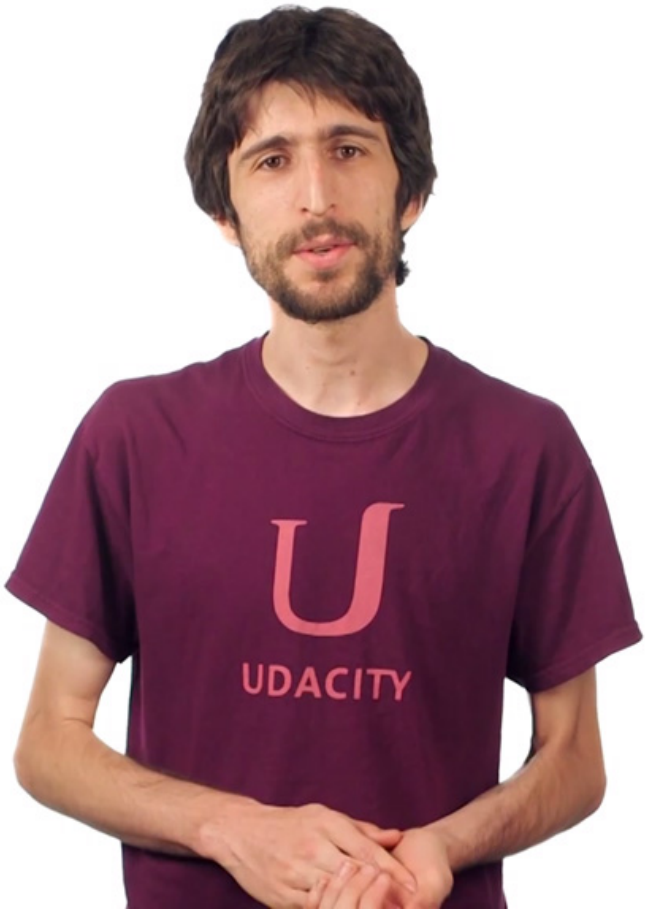
Each line should be less than 80 characters, and please add related instructions for your functions.

Good coding style is extremely important. Google Python Style Guide is a good reference.

Resubmit Resubmit Project

DOWNLOAD PROJECT

4 CODE REVIEW COMMENTS ›



Rate this review
( ) ( ) ( ) ( ) ( )

Best practices for your project resubmission

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

Watch Video (3:01)

Have a question about your review? Email us at review-support@udacity.com and include the link to this review.

Return to Path

Rate this review
( ) ( ) ( ) ( ) ( )

- Student FAQ