

[Return to Classroom](#)[DISCUSS ON STUDENT HUB](#)

# Machine Learning Capstone Project

REVIEW

CODE REVIEW

HISTORY

## Meets Specifications

Your work was amazing!

You have good stuff here, and I encourage you to keep improving your work to go on with your great trajectory. 🙌



Good luck with your next projects!

Wish you the best of luck in your future!

If you want to add me on [Linkedin \(Rafael Buck\)](#) feel free.

Proost!!! 🍷

## 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.

Incredible work with the opening section. You gave good starting paragraphs to outline the project and provided basic information about the problem domain 😊

In addition, I found your theme a very interesting field of application for machine learning, and you have correctly proposed a practical approach to solving the problem.

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

Very objective explanation and good choice of your algorithm metric. And it matched perfectly with your problem 😊

**Suggested:** As machine learning engineers, it is always important to justify why we choose a specific metric to evaluate the performance of our model. We need to explain why some metrics are more important than others to the problem we are analyzing.

- I have attached here a [reference](#), for your future works, that I found about several metrics used for evaluation of machine learning algorithms.
- Here a fascinating reference, that I got from another good reviewer, about [Choosing the Right Metric for Evaluating Machine Learning Models](#).
- And here, another one, that discusses [What metrics should be used for evaluating a model on an imbalanced data set](#), very cool also.

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.

The problem statement is clearly defined.

## 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.

The exploratory analysis of the data is very good. One thing we need to do at this stage of the project (as well as your future machine learning projects) is to present more business-driven insights.

**Suggested:** Here is a complete article on [various techniques of the data exploration process](#).

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

Perfect! I'm very impressed with the algorithms descriptions 😊

**Suggested:** It is always a good idea to provide links to references, images, diagrams and examples to validate the information we present in our work.

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, we're on the right track. Here the idea is, as a machine learning engineer, to introduce early insights to a more business audience, for example investors for your project, and you did very well 😊

Here an interesting reference that could be used in future works to explain some machine learning concepts to your readers: [Google's Machine Learning Glossary](#)

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

**Suggested:** Here is a link to a very cool article about the importance of the benchmark in machine learning work: <https://blog.dominodatalab.com/benchmarking-predictive-models/>

## Methodology

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.

Excellent step by step process. Your results would definitely be replicable, and you were very clear of how achieved them. Congrats!!!

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.

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 job documenting all your pre-processing steps 😊

In general (it always depends a little on the specific problem), it is still good in this step to check some characteristics of the variables in your dataset, such as identification and strategy with outliers, linearity between variables, balancing the dataset, etc.

## 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.

Amazing!

**Suggested:** I will leave here a very cool reference for us, machine learning engineers, to use to explain results. I consider this book sensational and very cool: <https://christophm.github.io/interpretable-ml-book/>.

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.

Very well documented the optimised model. Congratulations!

 [DOWNLOAD PROJECT](#)

[RETURN TO PATH](#)