UDACITY

PROJECT

# Machine Learning Capstone Project

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

## Requires Changes

9 SPECIFICATIONS REQUIRE CHANGES

Hi,
First of all, you have some really strong features in your project such as the usage of supervised learning combined with Q-learning and the data retrieval workflow/implementation. Even though, there are a few open requirements, they are not going to turn your project upside down, they are all confined to the report.
Finally, you do express concern about the document length, but in fact **you can go from 9 up to 15 pages (and not 12 pages)** for the capstone report, so you still have room to add more information to this report.

Best Wishes.

### 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.**

1. **required:** you need to provide a high level overview of project. For example, when you mention the following:

   > "The area of focus is Investment and Trading, narrowed to trading with SP500 stocks."

   You should provide after this statement a high-level explanation about what is the stock market and why it is relevant (and difficult) to solve problems in this domain.

**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.**

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

1. **required:** you need a section to describe and justify the metrics used in this project. For each metric that you use to evaluate your model you should explain why it is significant to use that metric for this problem. Also, if the metric is not well-known in the machine learning community you might want to include an example of how it is calculated along with its formula.
2. **suggestion:** create a new section for metrics, it may be part of Section I (as a subsection) or an entire section before section I.

### 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.**

1. **awesome:** good job on the data acquisition documentation, Figure 1 is really useful.
2. **suggestion:** maybe use black/gray color for the legends in the diagram of Figure 1 (white on green is kind of difficult to read).
3. **required:** explain what are Bollinger Bands and Bull-Bear Spread
4. **required:** you should add a thorough data analysis here. You do mention and corrects for the NaN values, however you should calculate some relevant statistics about the data.
5. **suggestion:** for example, a first simple statistic is to run the describe() method of your pandas dataframe and analyze the mean and standard deviation of data (before the minMaxScaler), this shall be useful even for your justification for the minMaxScaler (see the requirement 1 for my comments on the preprocessing steps).

**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.**

1. **required:** add a meaningful visualization that summarizes or extracts a relevant characteristic about your dataset.
2. **suggestion:** for example, calculate the features correlations (you don't need to do that to all stocks) and plot them in a "m x m" matrix (such that m = number of features), where each entry corresponds to the correlation between the features of the respective column and row (see example below).

| | f1 | f2 | f3 | f4 | f5 | f6 | f7 | f8 | f9 | f10 |
|---|---|---|---|---|---|---|---|---|---|---|
| f1 | NaN | 0.43311731 | 0.8770782 | 0.83613852 | 0.404782 | 0.87913552 | 0.41997405 | 0.36087428 | 0.78589491 | 0.36521271 |
| f2 | 0.43311731 | NaN | 0.5339228 | 0.53749362 | 0.42422732 | 0.53599095 | 0.57291523 | 0.441674 | 0.54510686 | 0.51650435 |
| f3 | 0.8770782 | 0.5339228 | NaN | 0.859593 | 0.38865267 | 0.85498413 | 0.41135347 | 0.35602409 | 0.81671936 | 0.37192468 |
| f4 | 0.83613852 | 0.53749362 | 0.859593 | NaN | 0.37947508 | 0.82905403 | 0.4056536 | 0.35097673 | 0.77770302 | 0.37260752 |
| f5 | 0.404782 | 0.42422732 | 0.38865267 | 0.37947508 | NaN | 0.52517275 | 0.56010568 | 0.64124111 | 0.56010568 | 0.74797275 |
| f6 | 0.87913552 | 0.53599095 | 0.85498413 | 0.82905403 | 0.52517275 | NaN | 0.4142441 | 0.36463311 | 0.78332215 | 0.35580379 |
| f7 | 0.41997405 | 0.57291523 | 0.41135347 | 0.4056536 | 0.56010568 | 0.4142441 | NaN | 0.3744452 | 0.56833127 | 0.39051618 |
| f8 | 0.36087428 | 0.441674 | 0.35602409 | 0.35097673 | 0.64124111 | 0.36463311 | 0.3744452 | NaN | 0.41851292 | 0.36600362 |
| f9 | 0.78589491 | 0.54510686 | 0.81671936 | 0.77770302 | 0.56010568 | 0.78332215 | 0.56833127 | 0.41851292 | NaN | 0.38757696 |
| f10 | 0.36521271 | 0.51650435 | 0.37192468 | 0.37260752 | 0.74797275 | 0.35580379 | 0.39051618 | 0.36600362 | 0.38757696 | NaN |

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

1. **required:** you only mention the techniques PCA, SVM, Random Forests, Decision Trees and kNN, AdaBoost, NaiveBayes and Q-Learning. However, you need to include a thoroughly discussion followed by a reasonable justification for each technique used based on the problem characteristics.
2. **suggestion:** for example, to justify using PCA and assuming you have analyzed the features correlation you may explain that you have used PCA to reduce the amount of redundant features.

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

**required:** you use the index in section III as your benchmark, however you should also clearly explain how the index portfolio was developed and the overall results that it produced.

"SIMULATION RESULTS INDICATE PERFORMANCE ABOVE THE INDEX, WHICH WAS OUT BENCHMARK"

**suggestion:** maybe it is for the best if you create a subsection within section III specifically for explaining your benchmark.

## 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.**

1. **required:** add an justification of why you use the MinMaxScaler for each feature.

   "IN THE NEXT STEP SCIKIT'S MINMAXSCALER TRANSFORMS MOMENTUM, SMA, MIN, [...]"

**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.**

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.

required: currently you only report a final Q-learning model with fixed parameters. You can try to optimize these parameters to obtain a refined model. I would not advise trying to tune the multitude of learning algorithms used to predict the labels though (since there are too many), but you may try if you want.

## 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.

required: you need a more reasonable justification for the parameters alpha and gamma. You may try the values that obtained good results in the smartcab project, however you cannot use that as a justification in this project.

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.

awesome: you did a thoughtful analysis in section VI, well done.

## Conclusion

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

1. awesome: Figure 9 clearly depicts the better performance of your approach in comparison to the benchmark.
2. suggestion: enlarge Figure 9, it is quite difficult to evaluate it without zooming in the document.

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

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.

awesome: you section IV discuss some interesting approaches to make the whole implementation more realistic.

## 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.

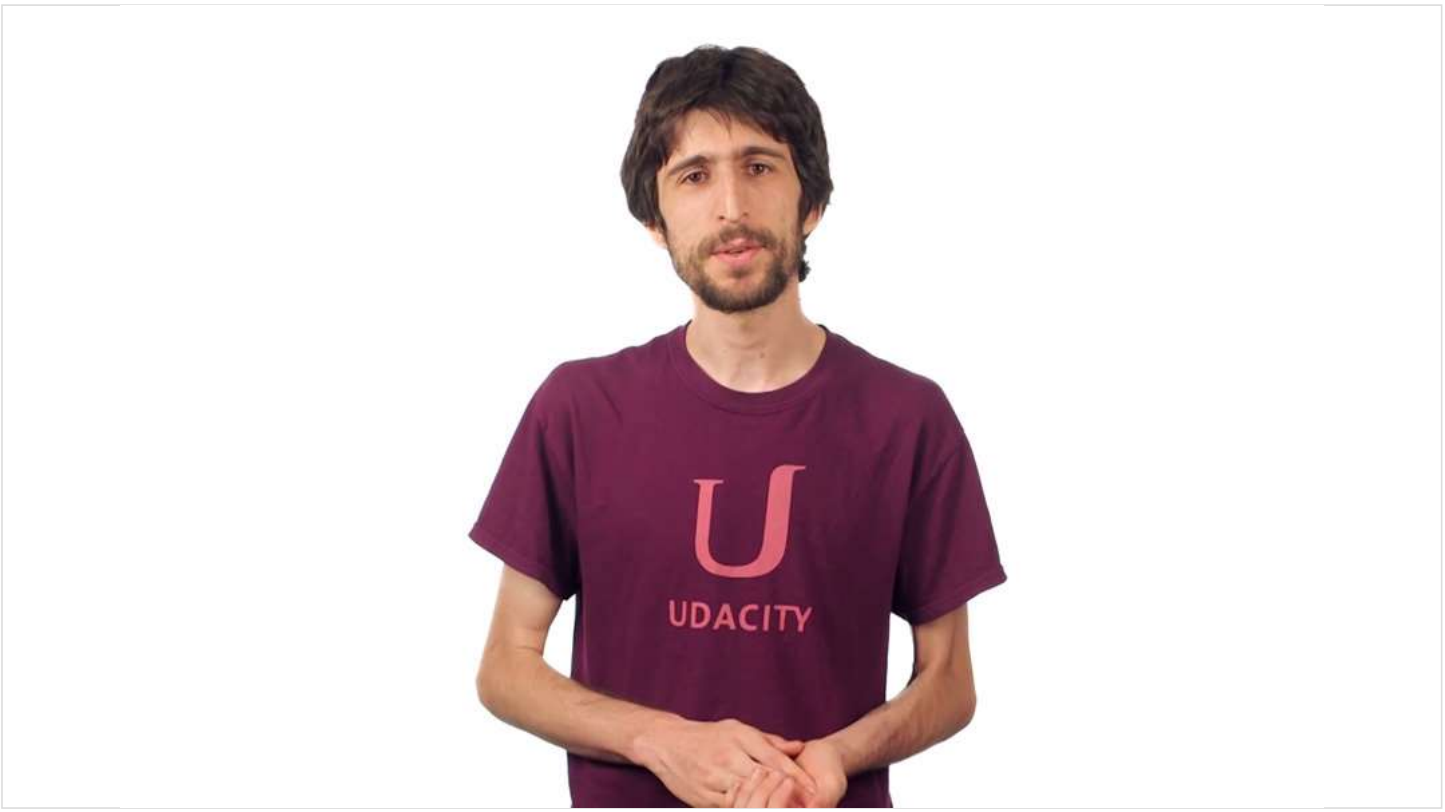
suggestion: there are multiple opportunities for you to combine separate phrases into paragraphs. For example, in page 2 you could combine the last 2 paragraphs (1 phrase each) with the previous paragraph ("The results are being...").

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

awesome: your readme.md is really useful, also your code is reasonably commented which is a good thing.

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)

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