



[< Back to Machine Learning Engineer Nanodegree](#)


Machine Learning Capstone Project

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

 Congrats on completing the capstone project! [Deep learning](#) is quickly becoming [software 2.0](#), so it's great that you're getting more practice with using it for text classification problems like this.

And if this is your final step to complete the MLND program, congratulations and best of luck with whatever projects you work on next! 

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.

Good introduction to the project and providing background information on the twitter sentiment analysis problem domain.

To see the increasing importance of solving prediction problems like this with machine learning, just look at how many ML APIs are now available:

- <https://www.kdnuggets.com/2018/05/50-useful-machine-learning-prediction-apis-2018-edition.html>

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.

Great work describing the metrics used to evaluate the model's performance, although technically the writeup should also show the actual equation(s) used to calculate the scores. 😊

For thoughts on other [metrics](#) to use whenever the target classes are imbalanced, you can also check out this [forum post](#) on the [Matthews Correlation Coefficient](#).

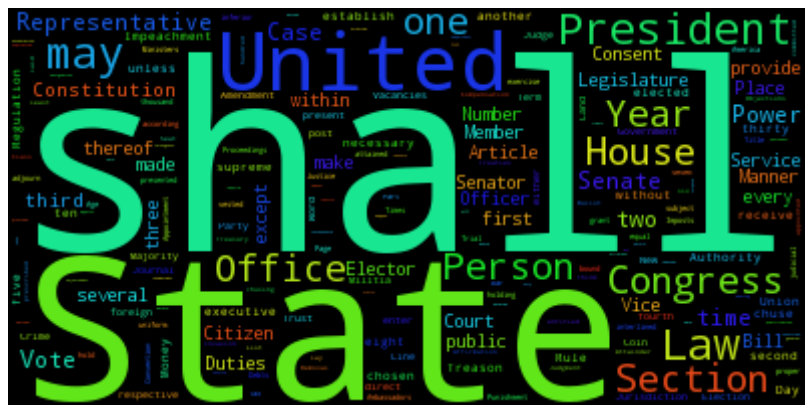
Analysis

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.

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 job discussing the dataset, and including the histogram [visualization](#) of the tokens per sentence to help readers see some important characteristics of your model inputs. 😎

If interested, you could also think about creating a [word cloud](#).



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

Good job including a discussion of the NLP and CNN modeling techniques used. Kudos!

You may have already looked into these, but here are some [NLP](#) resources that might be worth exploring:

- <http://ruder.io/deep-learning-nlp-best-practices/index.html>
- <https://explosion.ai/blog/deep-learning-formula-nlp>
- <https://www.kaggle.com/getting-started/55298>

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

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.

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.

Great job describing the implementation of the CNN model and documenting refinements you made with the tokenization. You definitely put in some work here to generate a good solution. 😊

- This section would also be a good place to describe if any complications arose. Did you have to adjust your expectations going into the project to find a good solution?
- In general it's also a good idea to record the results of all your model training experiments, especially "failures" that might need to be debugged. You can learn more about this and other thoughts from google in their [effective ML guidelines](#).

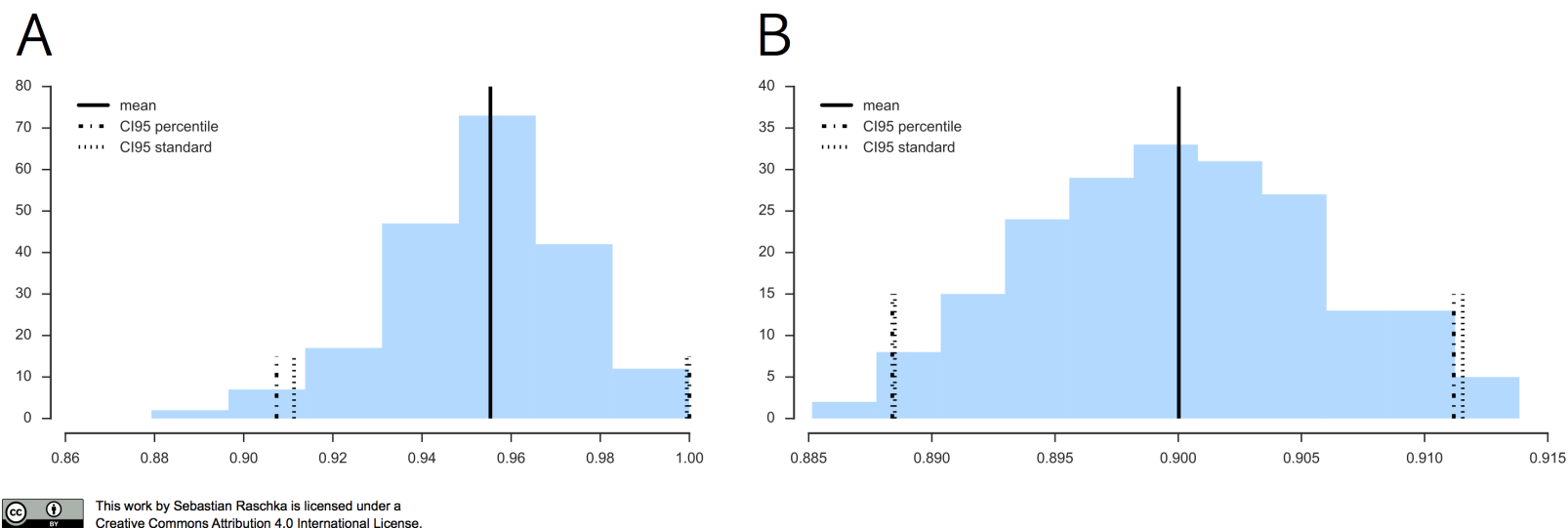
The process of improving upon the algorithms and techniques used is clearly documented. Both the initial and final solutions are reported, along with intermediate solutions, if necessary.

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.

Good job discussing your final models results, and congrats on meeting your benchmark. 😎

You can also consider taking steps to [validate your models](#) by calculating [bootstrap confidence intervals](#).



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.

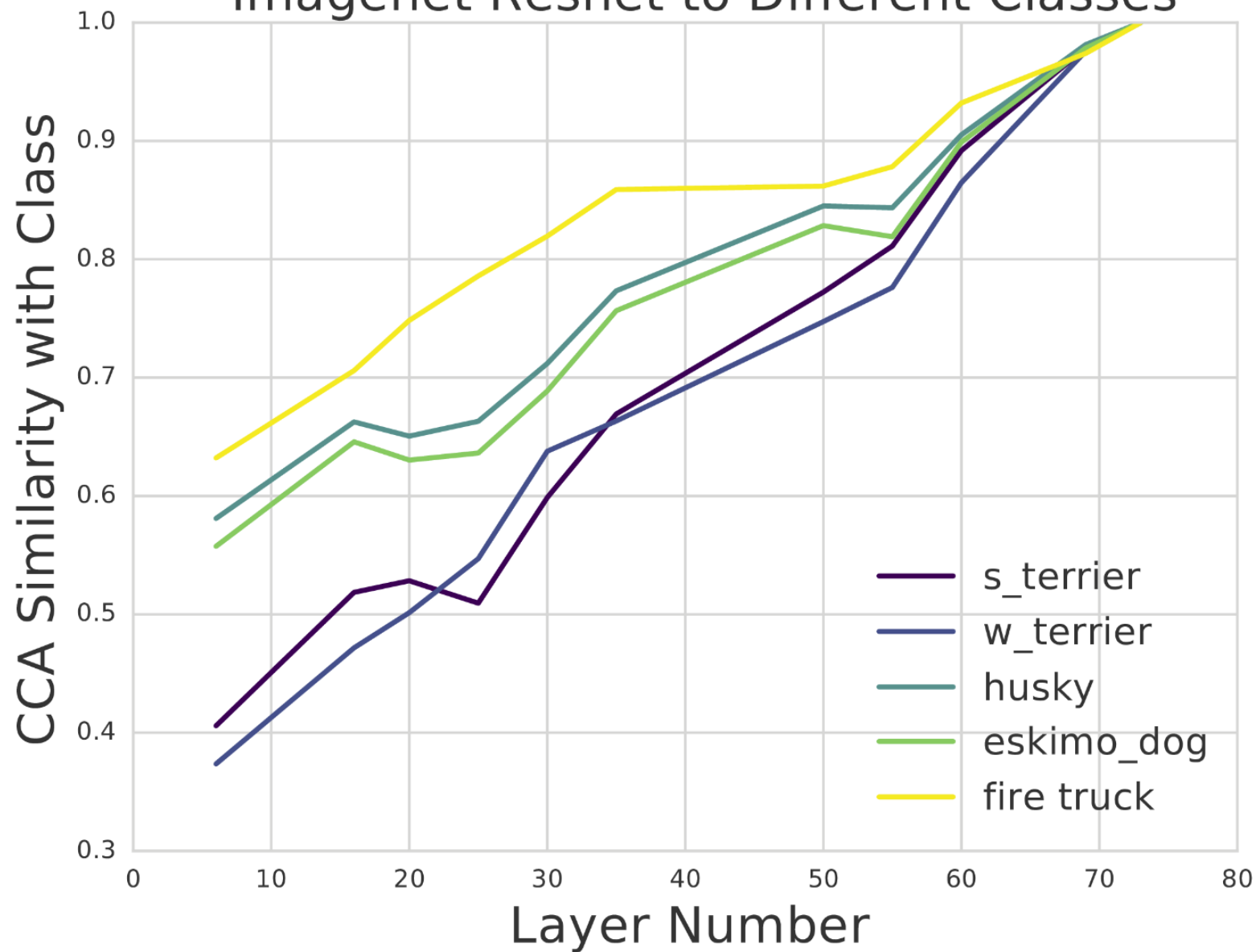
Conclusion

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

Nice visualization of the confusion matrix to illustrate how your model is learning to predict the data. Well done!

If you want to interpret neural network models, you can also try measuring the similarity of different model layers to the target classes, as described in [this research by Google](#):

CCA Similarity (using DFT) of Layers in Imagenet Resnet to Different Classes



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 reflection on your process for the project and discussion of improvements that could be made with more data and using a different RNN model architecture.

You've clearly expanded your skills beyond the core concepts introduced in the MLND and are ready tackle new NLP problems like this with [machine learning](#). 😊

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.

If you haven't already used them yet, it could be cool to try out one of the newer user-friendly GPU cloud services: [Floyd](#), [Crestle](#), [Paperspace](#), [Vectordash](#), [Clouderizer](#), [snark.ai](#).

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.

 [DOWNLOAD PROJECT](#)

RETURN TO PATH

[Student FAQ](#)