UDACITY

PROJECT

Machine Learning Capstone Project

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

🏆 Congrats on completing the capstone project! I'm impressed with how the results of your stacking of various regressors turned out.

And if this is your final step to complete the MLND program, best of luck to you on whatever projects you decide to 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 work outlining the project and providing background information on the housing price problem domain. This is definitely a real world problem that can be tackled with machine learning. 😄

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.

Nice job defining the regression prediction problem, and discussing how the solution needs to predict the Ames housing price data.

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

Good work describing the Log RMSE metric you'll use to evaluate the model's performance.

You could also think about using Mean Absolute Percent Error (MAPE) to scale for the differences in selling prices across different types of houses.

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.

Good discussion of the Ames dataset, including some helpful summary stats and samples of what the dataset contains — this is an important part of the analysis and helps readers understand just how tricky a problem this is. 😉

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 set of visualizations that help us explore the data and see what kind of sale prices and feature correlations we're dealing with. 😎

For future visualization ideas, you can also check out this new visualization tool released by Google.

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 here of the regression algorithms and stacking techniques used and why you chose them. Kudos!

Another simple sklearn model you could try out here is an MLP regressor.

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

Nice work providing a reasonable benchmark for the model using the kaggle score. 😎

To simplify the benchmark, you could also try a naive predictor that always predicts the mean or median housing sales price.

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.

Good documentation of your data processing steps, including creating the train/test splits and steps performed in transforming and engineering the input features.

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 and documenting your training times and model scores.

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.

Nice use of grid search to improve the models' performance and then using stacking to create a more powerful meta-learner. 😅

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 model and validating the robustness of the stacked average model by submitting the results to kaggle. 😉

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.

Excellent work discussing the significance of the results, and congrats on meeting both the RMSLE and R2 benchmarks!

Conclusion

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

Very nice visualization that illustrates how the model is predicting the sales prices quite well at both lower and higher price ranges. Well done!

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 summarization of your process for the project and discussion of the feature engineering/selection and stacking aspects of the implementation.

You've clearly grasped the concepts introduced in the MLND while expanded your skills further into tackling new regression problems 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.

Good discussion of how future improvements could be made with things like more fine tuning of the features and models used.

It might be cool to also try a deep learning approach and use a GPU-enabled cloud resource like Floyd, Crestle, or Paperspace.
(a bit more user-friendly than using AWS) 😉

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