

Report: Predict Bike Sharing Demand with AutoGluon Solution

BANALA DEEPTHI

Initial Training

- 1) What did you realize when you tried to submit your predictions? What changes were needed to the output of the predictor to submit your results?

TODO:

The model did not perform as well as expected the first time I used the raw dataset without performing any data analysis or feature engineering. I discovered that the prediction had some negative RMSE score values while I was submitting it. We now know that if we don't set everything to be bigger than 0, Kaggle will reject the prediction. To deal with this situation, I first counted the number of negative score predictions the predictor returned before setting all of the negative predictions to 0. I assign those predictions to the submission dataframe at the end.

- 2) What was the top ranked model that performed?

TODO:

The WeightedEnsemble_L3 model was the top ranked model which performed well with the created features.

Exploratory data analysis and feature creation

- 3) What did the exploratory analysis find and how did you add additional features?

TODO:

I separated the datetime into year, month, and hour characteristics for the added features. Each record has information about the time element because of these qualities. The transformation of the season and weather variables into categorical features with discrete values, in contrast to the rest of the features, which represent continuous data distribution, was also very helpful.

- 4) How much better did your model perform after adding additional features and why do you think that is?

TODO:

After the additional features were added, the model considerably outperformed itself. The RMSE score decreased from 1.78450 to 0.66172, showing an improvement.

Hyper parameter tuning

- 5) How much better did your model perform after trying different hyper parameters?

TODO:

Some configurations were helpful, but some other configurations degrade the model's performance. Hyper parameter adjustment was helpful in some situations, but it did not significantly enhance performance.

- 6) If you were given more time with this dataset, where do you think you would spend more time?

TODO:

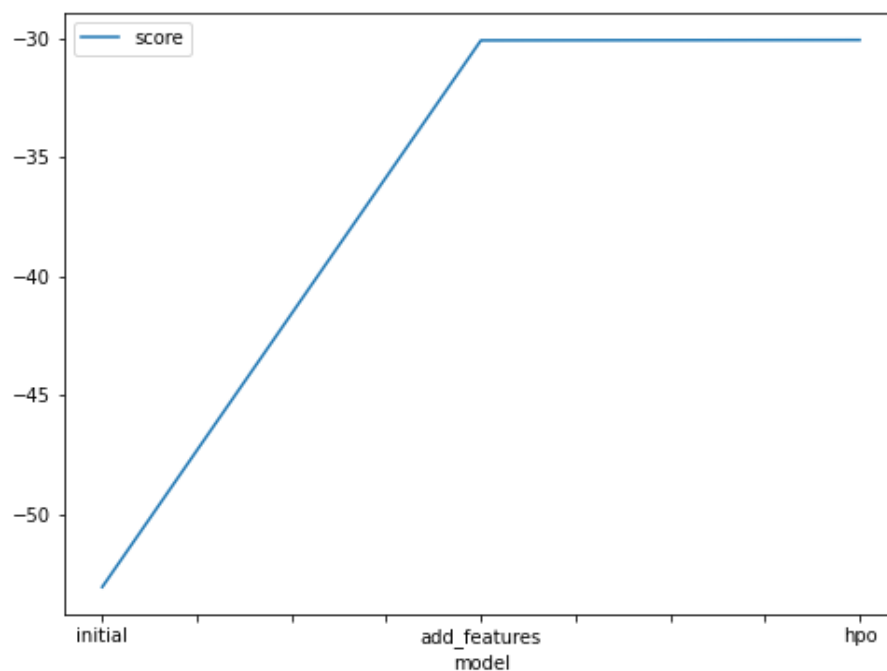
To learn more about this dataset, I shall do a more thorough data analysis. Choose the model that performs the best after evaluating many models. I'll optimise the hyperparameters of the best available model. Finally, I'd like to investigate various feature engineering strategies for enhancement.

- 7) Create a table with the models you ran, the hyperparameters modified, and the kaggle score.

model	hpo1	hpo2	hpo3	score
initial	8	3	25	1.78450
add_features	8	3	25	0.66172
hpo	5	5	25	0.69751

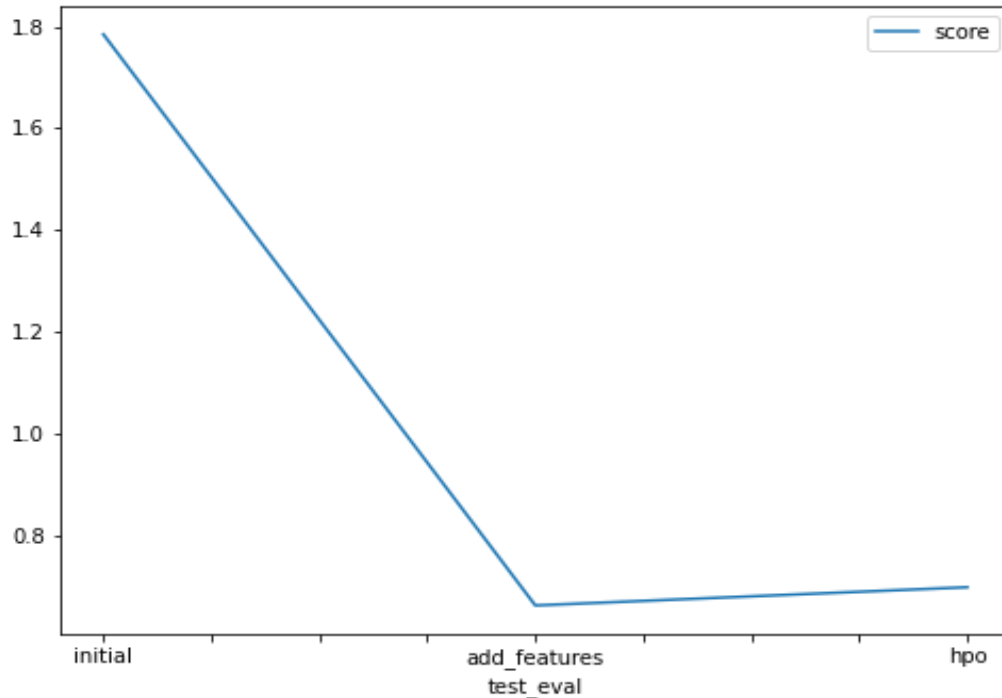
- 8) Create a line plot showing the top model score for the three (or more) training runs during the project.

TODO:



- 9) Create a line plot showing the top kaggle score for the three (or more) prediction submissions during the project.

TODO:



10) Summary

TODO:

After the training of the Bike Sharing demand dataset, I created the original first model using the default parameters without messing with the function too much. I received my first RMSE score after model training. We then trained the model on test data and the RMSE Kaggle score improved significantly. I decided to fine-tune the model algorithm's hyperparameters using the preprocessed features because the RMSE score increased somewhat, from 0.66172 to 0.69751. The Autogluon framework's training on the demand data for bike sharing demand produced positive outcomes in the end.