Report: Predict Bike Sharing Demand with AutoGluon Solution

Martin Barrientos

Initial Training

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?

I realized that the submission got rejected if any of the predictions were negative. In order to fix that, I had to set any negative values equal to zero.

What was the top ranked model that performed?

The top ranked model was WeightedEnsemble_L3.

Exploratory data analysis and feature creation

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

The exploratory analysis showed the distribution of each feature relative to the data. I added additional features by parsing the datetime strings into datetime objects. I, then, parsed the year, month, day, and hour. I, also, set the dtype to category for the “season” and “weather” features.

How much better did your model preform after adding additional features and why do you think that is?

After adding additional features, the model train score improved from -115.161955 to -35.469634, and the kaggle score improved from 1.39257 to 0.45676.

Hyper parameter tuning

How much better did your model preform after trying different hyper parameters?

After trying different hyper parameters, the model train score went from -35.469634 to -36.105422. The kaggle score went from 0.45676 to 0.45408.

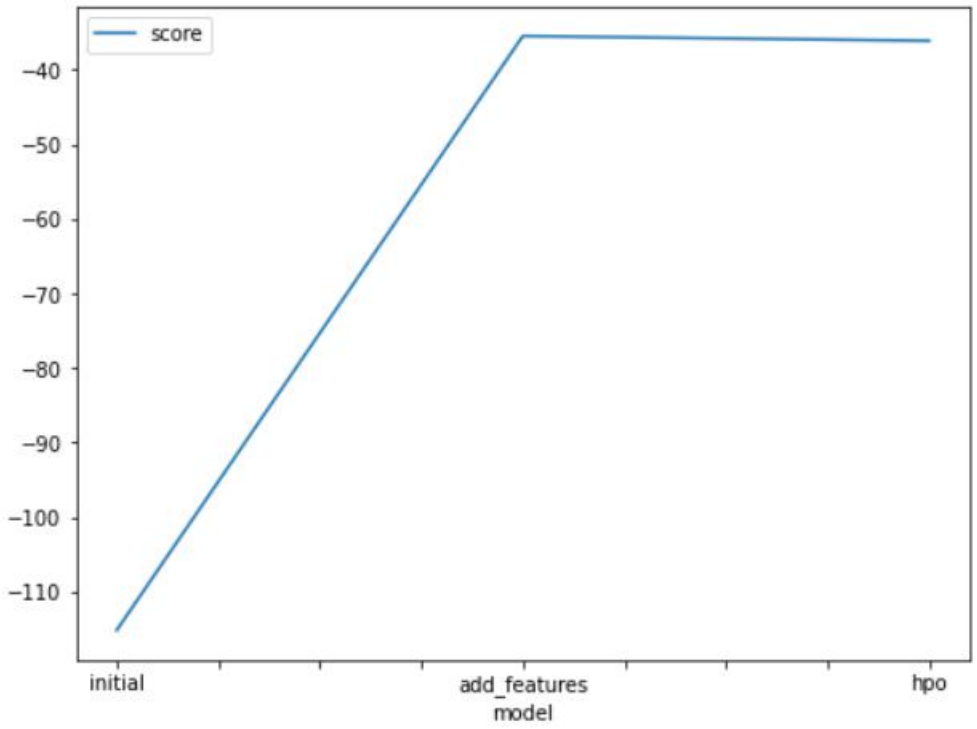
If you were given more time with this dataset, where do you think you would spend more time?

If I were given more time with this dataset, I would start by creating more features because that seemed to give a greater improvement than attempting to optimize the hyperparameters. Given more time, I would also attempt to fine tune more hyperparameters as well as adding more training time.

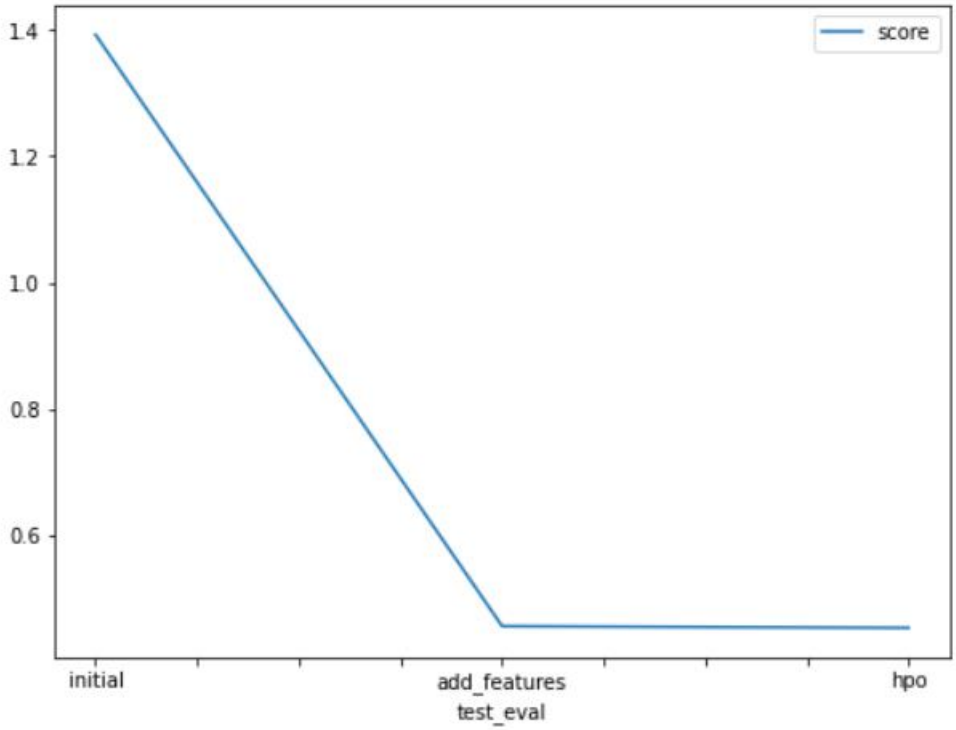
Create a table with the models you ran, the hyperparameters modified, and the kaggle score.

model	time	num_bag_folds	num_stack_levels	score
initial	600	default vals	default vals	1.39257
add_features	600	default vals	default vals	0.45676
hpo	900	5	2	0.45408

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



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



Summary

I started out by training my model with the default parameters to get a good basis for comparison. I, then, created new features and changed numerical feature values to categorical dtypes. Lastly, I fine tuned some hyperparameters in order to further improve the model's score. In the end, the best performance was achieved by creating new features and fine tuning some hyperparameters.