Report: Predict Bike Sharing Demand with AutoGluon Solution

Maryam Alqarrous

1. Initial Training

1.1. 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?

The initial submission went successfully because the project template contains all the requirement, so I did not have to change anything to submit my result.

1.2. What was the top ranked model that performed?

The top ranked model was the hyper-parameterized achieving 0.48847 of Kaggle score.

2. Exploratory data analysis and feature creation

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

The exploratory analysis found that the datetime column can be split into hour, day and month which might enhance the model understanding about the time of the day, the day of the month or the month in a year and how it affect the target variable which is count. Also, there was some int column that in reality represents categorical data such as the season, weather, holiday and working day, butter to be converted into categorical. Overall, adding additional features affected the Kaggle score by reducing it from the 1.84171 to 0.61717.

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

Adding additional features affected the Kaggle score by reducing it from the 1.84171 to 0.61717. I think that adding additional features enhanced the model understanding and helped it find more patterns in the data.

3. Hyper parameter tuning

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

After adding additional features, the model achieved 0.61717, and after trying different hyper parameters it achieved 0.48847.

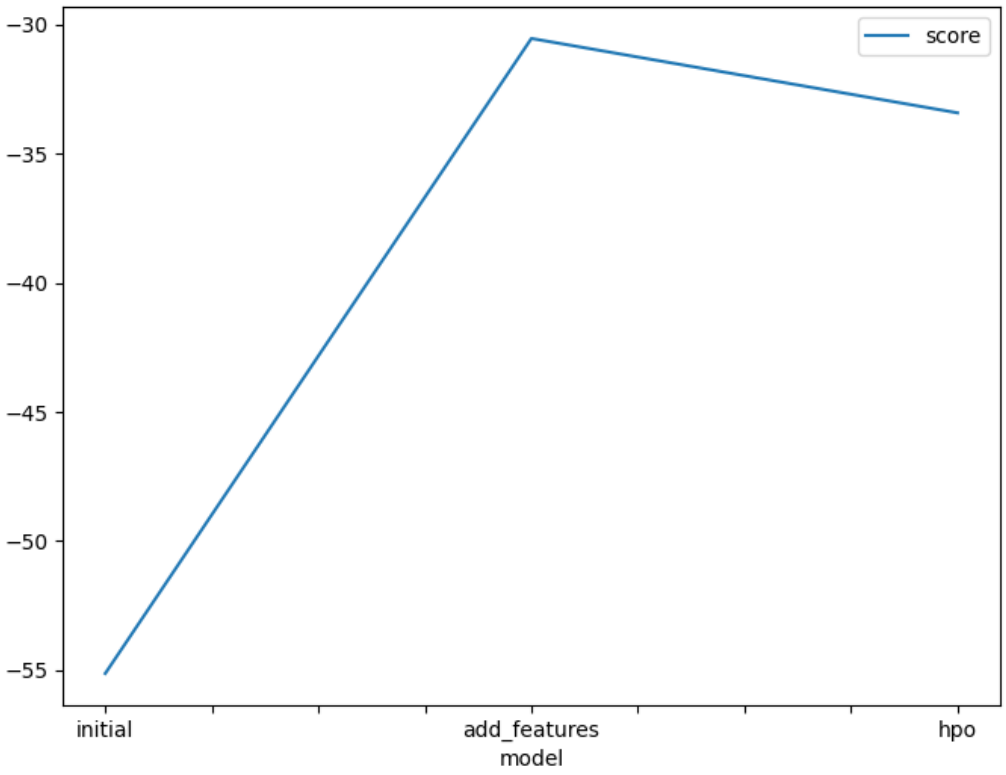
3.2. If you were given more time with this dataset, where do you think you would spend more time?

I would spend more time on tuning the model with that achieved high performance and try to enhance it more.

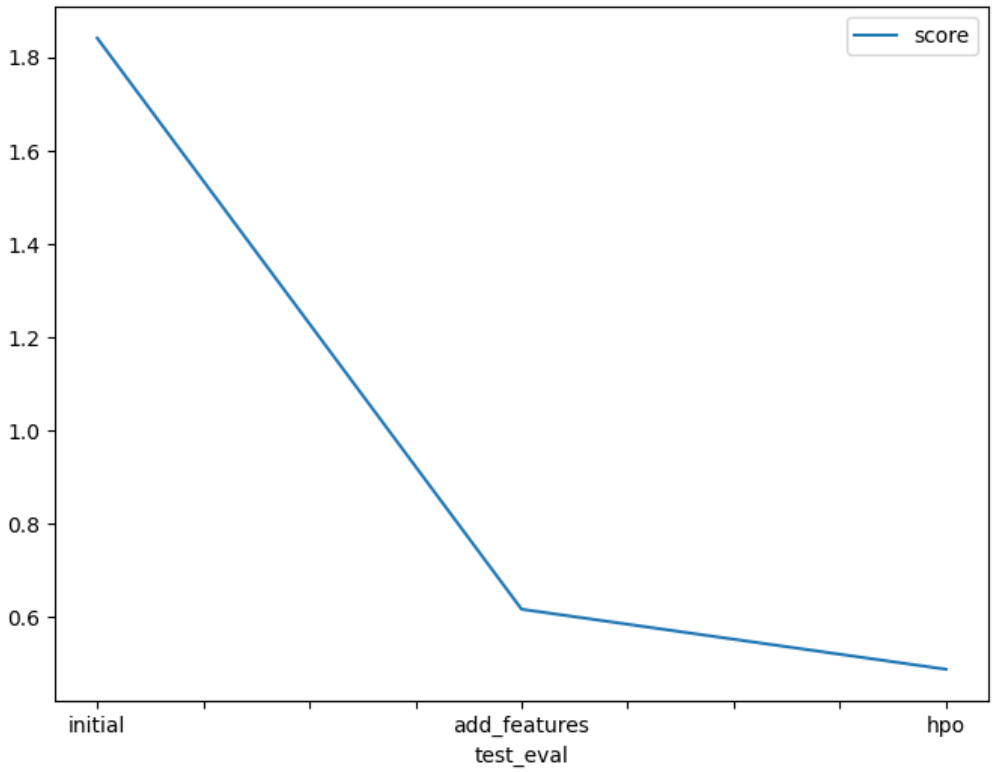
3.3. Create a table with the models you ran, the hyperparameters modified, and the Kaggle score.

	model	time	eval_metric	hyperparameter	score
0	initial	600	root_mean_squared_error	default	1.84171
1	add_features	600	root_mean_squared_error	default	0.61717
2	hpo	600	root_mean_squared_error	'num_boost_round':500, 'n_estimators':500, 'se...	0.48847

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



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



4. Summary

Overall, this experiment shows that the AutoGluon is useful and easy to use specially with the problems that needs to be trained on different model to decide which model is the best for the problem without the need to have extensive knowledge of all the models. But it requires some knowledge on the hyper parameter tuning and data engineering to get the best results.