Report: Predict Bike Sharing Demand with AutoGluon Solution

SALVATION PETER

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?

First, I needed to check for negative values in the prediction. Some of the times I trained the model, I had negative values. Then, I changed all the negative values to 0 as negative values returned errors in the submission.

Secondly, I ensured the datetime column in the submission file is the same as the datetime in the test dataset, then I changed the count column to reflect the predictions from the test dataset. After this was completed, I submitted the first time.

What was the top ranked model that performed?

The top-ranking model at this point is the WeightedEnsemble_L3 with a score of -53.086669

Exploratory data analysis and feature creation

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

In the exploratory analysis, I observed that the seasons, weather and workingday columns were plotted as numbers rather than categories.

Also, some columns were not normally distributed.

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

My model performed 2 times better than before with the addition of more features giving a new score of 0.67527

Hyper parameter tuning

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

There was not much improvement from the last predictions compared to the improvements from the 1st model.

In actual fact, the performance of my model with the hyperparameters on the predictions gave a score of 0.51132 even though the best performing model was the WeightedEnsemble_L2 with a score of -32.184271

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

I would ensure the columns were normally distributed such as the windspeed and the humidity columns. I learnt some ML algorithms do better with normally distributed features.

Also, I would do a correlation matrix to ensure the features I train the model with actually correlate with the target feature.

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

S/No.	model	presets	problem_type	time_limit	score
0	initial	best_quality	N/A	600	1.79657
1	add_features	best_quality	regression	600	0.66308
2	hpo	optimize_for_deployment	regression	720	0.51132

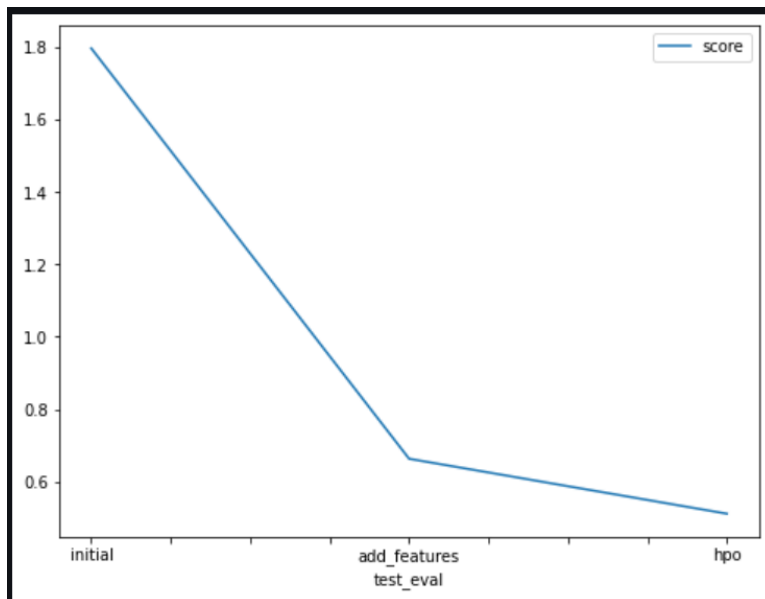**Create a line plot showing the top model score for the three (or more) training runs during the project.**

![model_train_score.png](img/my_model_train_score.png)
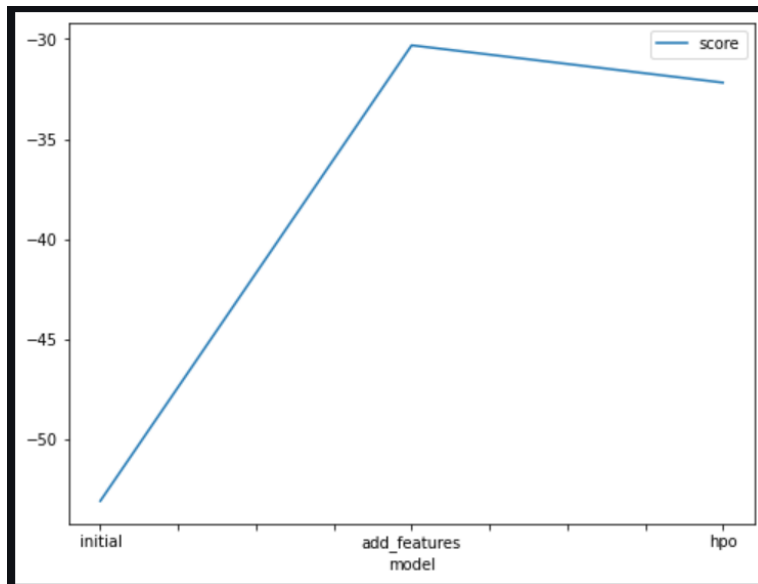
Create a line plot showing the top kaggle score for the three (or more) prediction submissions during the project.

![model_test_score.png](img/model_test_score.png)

Summary

In summary, after I had tried optimizing in different ways, I can say that the tunning of hyperparameters are great in improving models as well as their accuracy scores in prediction but it is not always the case. This is because knowing the right hyperparameters to tune as well as having a good understanding of the features in the dataset is of great importance as these greatly affect output.

I would also add that it is best to begin training a model without the hyperparameters so that when one begins to adjust the parameters, he can easily tell if there is an improvement or not and would be better equipped to handle them.

Also, Autogluon is a good tool as it helps test several models in one go.