Report: Predict Bike Sharing Demand with AutoGluon Solution

Sara Ahmed Maher

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?

All values in the data frame must be positive, using the .descripe() function we can find some values are negative.
I changed all negative values to zeros so that kaggle submission accepts them.

What was the top ranked model that performed?

The best kaggle score i got was the second submission with new features.

Exploratory data analysis and feature creation

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

Plotting the histogram of all features showed what the data represents, like categorical or binary or continuous representations.

I did split the datetime feature into different features adding weight to each individual detail in the datetime.

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

RMSE scored 0.44182 while initially having a score of 1.77783 which is significant improvement.
This was expected as additional features increase the complexity of the model. Which in turn can increase the accuracy of the model.

Hyper parameter tuning

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

Trying different hyper parameters also resulted in good scores.
It was significant change compared to the initial score.

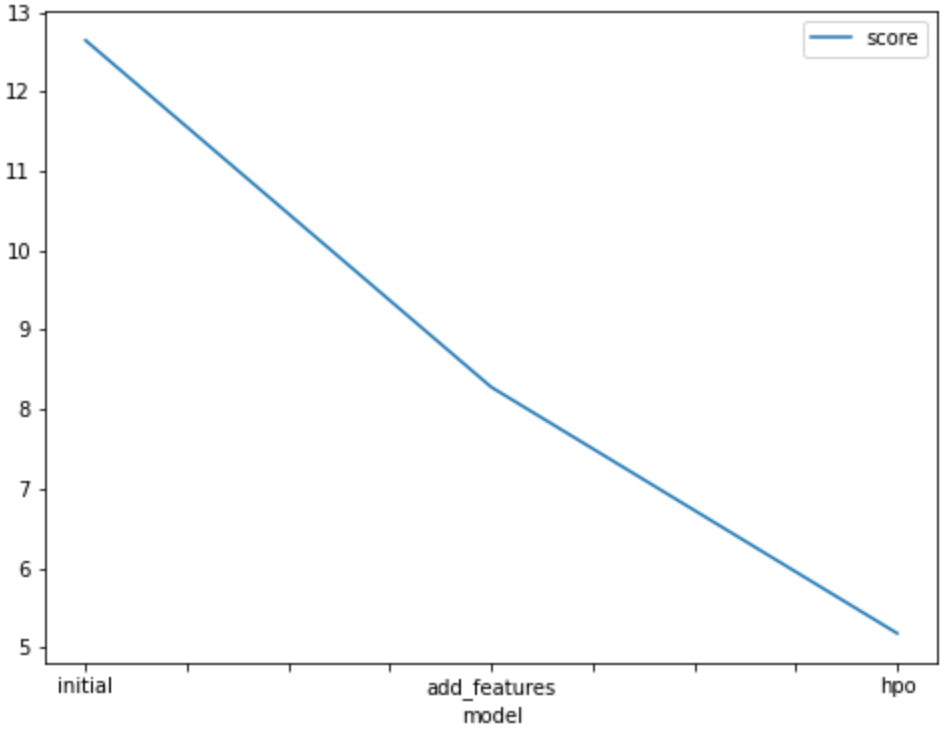
If you were given more time with this dataset, where do you think you would spend more time?

With this dataset first i would try to look at different data processing techniques
Then i would try new algorithms other than autogluon, then based on which model performs the best i would try hyperparameter tuning on it.

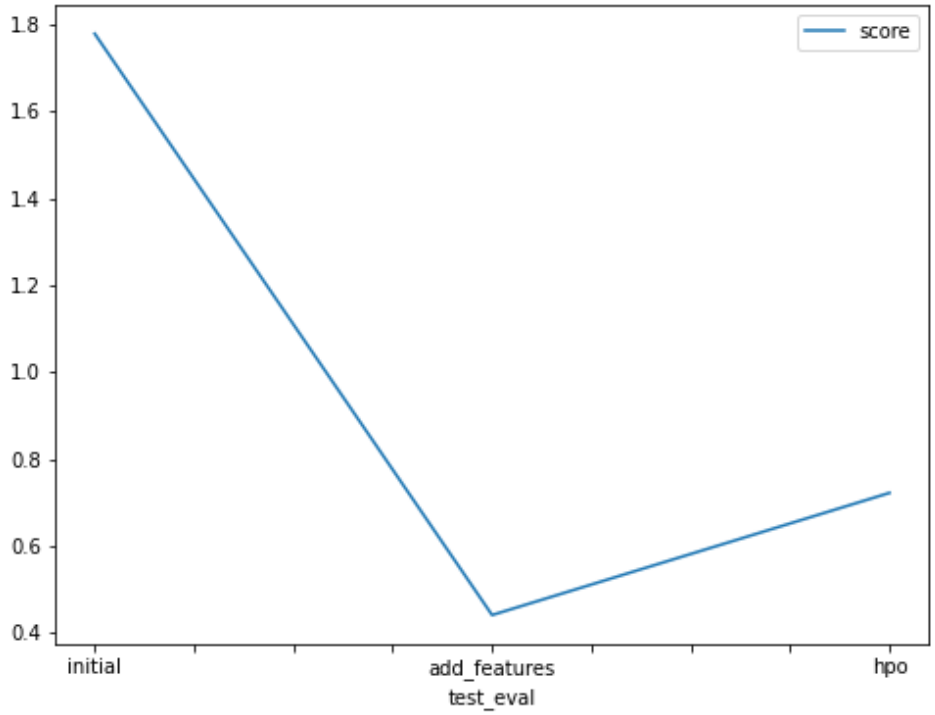
Create a table with the models you ran, the hyperparameters modified, and the kaggle score.

	model	presets	time_limit	num_bag_folds	score
0	initial	high_quality	600	-	0.51568
1	add_features	best_quality	600	5	0.44475
2	hpo	best_quality	800	5	0.49897

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


13
12
11
10
9
8
7
6
5
score
initial
add_features
hpo
model

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



Summary

The goal of this project was to build a regression model to predict bikes rented count using AutoMl.
The initial step was to run the mode with the dataset as is which didnt give very good results.
Afterwards Exploratory data analysis was used, observing their results we made two changes first added additional features and second we notice that two columns are treated as normal integers when they are categories so we change there data types.
The we rerun the model after these changes and notice a significant improvement in the score.
Afterwards we try different hyperparameters to see how the results are affected