Report: Predict Bike Sharing Demand with AutoGluon Solution

Desmond Osarfo Amoah

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?

Although the predictions generated were not exactly as expected, there were no negative values within the predictions. This makes sense as count is always greater than or equal to zero. No errors were generted when i tried to submit my predictions as a result.

What was the top ranked model that performed?

My top ranked model was WeightedEnsemble_L3 after the initial (1st) training.

Exploratory data analysis and feature creation

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

Histograms generated from the dataset revealed that:

1. There are four categories seasons
2. three categories of weather Also by describing the train and test dataset, it was obvious that some categorical features are object types (str types) in the dataset. This might somehow influence the decisions of the model. So categorical features like the season and whether were converted to type: categorical. Additional features like the hour, day, month and year were extracted and added to the dataframes.

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

The score value of the model improved considerably to a value of 0.67673. This imporved score can be attributed to the introduction of new features like day, time, month, year and changing categorical datatypes to their appropriate type. This has helped better the predictive performace of the model.

Hyper parameter tuning

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

I mut say i was amazed at how the score value improved further by introdutiocn of some new hyperparameters. the sore value reduced to 0.48059. which was the lowest so far, An improvement of the previous model.

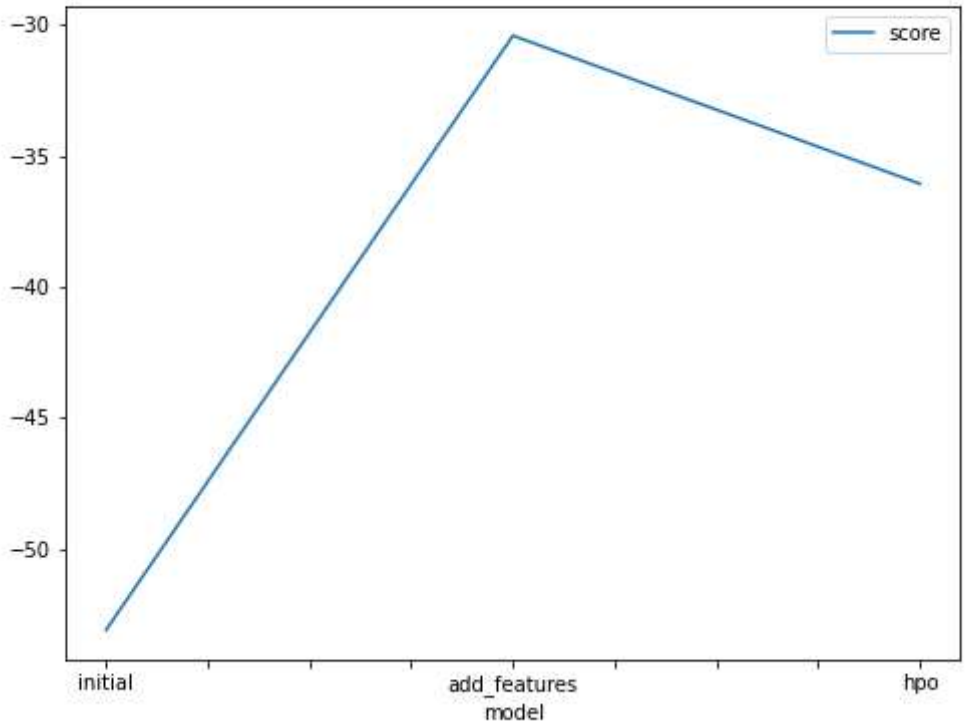
If you were given more time with this dataset, where do you think you would spend more time?

Given more time I might to consider feature engineering other columns of the data like, temp, atemp, humidity and others.

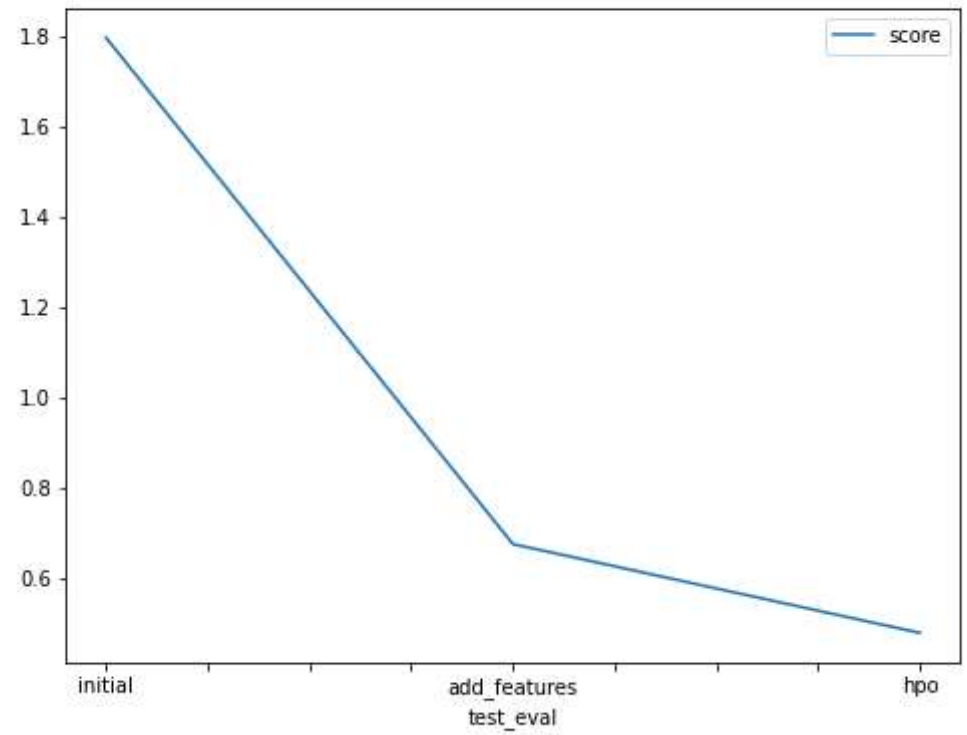
Create a table with the models you ran, the hyperparameters modified, and the kaggle score.

model	hpo1	hpo2
initial	eval_metric="root_mean_squared_error"	-
add_features	eval_metric="root_mean_squared_error"	problem_type = "regression"
hpo	eval_metric="root_mean_squared_error"	hyperparameter_tune_kwargs={'scheduler' : 'local','searcher': search_strateg

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

A regression model was built from the Kaggle bike sharing demand data using AutoGluon. Feature engineering and Hyperparameter optimization turned out to improve the models performace considerably.

NB: the images for the plots have been included in the folder incaes they are not showing.