Report: Predict Bike Sharing Demand with AutoGluon Solution

Aya Hesham

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?

The model performance improves as we do relevant changes like feature engineering or hyperparameter tuning.

What was the top ranked model that performed?

WeightedEnsemble_L3 with hyperparameter optimization

Exploratory data analysis and feature creation

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

- Looking at correlation matrix, the target label is highly correlated with "temp" and "hour" columns which totally makes sense as these two factors are very important to any person to decide whether to ride a bike or not.
- The demand changes monthly and yearly as well.

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

The model improved by 65.6% as adding specific features direct to the model can help improve model performance.

Hyper parameter tuning

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

the rmse value decreased by 25% which is good improvement.

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

I'd explore the relationship between each feature and the target label using bi-variate plots like scatter plots.
I might try one-hot-encoding categorical features.

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

	model	timelimit	presets	num_boost_round	score
0	initial	time_limit = 600	presets='best_quality'	100	172.29
1	add_features	time_limit=600	presets='best_quality'	100	59.03
2	hpo	time_limit=600	presets='best_quality'	150	44.05

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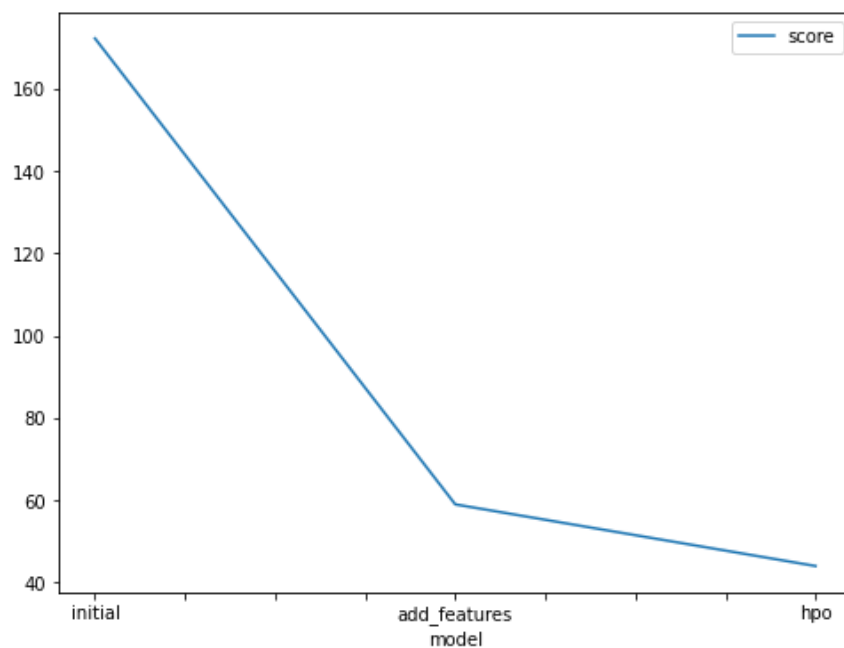
Create a line plot showing the top kaggle score for the three (or more) prediction submissions during the project.
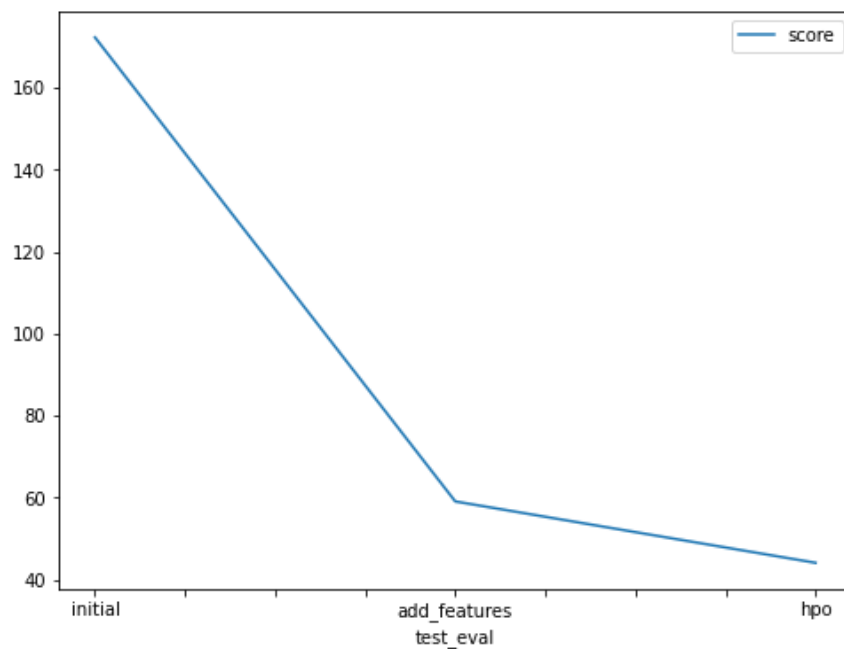
Summary

In this project bike demanding dataset was explored and ML lifecycle process were done

- Problem from a Kaggle competition to predict the bike-riding demand with autoML algorithm: AutoGluon
- The dataset was explored and analyzed with visualizations
- 3 different models of AutoGluon were made
    1. The default one
    2. After feature engineering
    3. Hyperparameter-optimized model
- For each model created, it was tested by a kaggle submission.
- This problem has a great business objective as it is related to multiple businesses like Uber.