Report: Predict Bike Sharing Demand with AutoGluon Solution

NAME HERE : Pushpita Das

Initial Training

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?

TODO: Before submitting my predictions I checked with the minimum value of the predictions and since it was non – negative , hence , no changes were there to be made . But still I checked with the negative values if there were any and since there were no such values I submitted my predictions successfully.

2. What was the top ranked model that performed?

TODO: The top ranked model came after hyperparameter optimization . The .fit() consists of the parameter named as “hyperparameters” it can be used to change the type and the number of models we want to train with, so, I used only RandomForest model to train my data, the RandomForest model has its own params for regression type problems which I used here.

Exploratory data analysis and feature creation

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

TODO: From the exploratory data analysis it was found that there was not much change in the demands across different seasons, also there is an uniform distribution across hours .

From the EDA the distribution for the column datetime was not accurate as the column included different months,years,days,hours,minutes and seconds so differentiating between them was difficult,hence, I extracted the months,years,days,hours,minutes and seconds columns from the datetime column and added to the dataframe.

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

TODO: It performed almost well after adding the additional features . Since I have used Root Mean Squared Error (RMSE) as the metric to evaluate the Performance so for the Initial model it came as -2.200765 But after adding the features it came as -30.537208.

So , the lower the RMSE the better the model hence the second model performed better.

Hyper parameter tuning

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

TODO: After trying different hyperparameters I realized the model performed more good than the previous ones.

After adding new features the Kaggle score was 0.61256 and after hyperparameter tuning the score became 0.76276.

so, the percentage change or increase is 15.02% , which itself show that tuning hyperparameters correctly can bring more improvement to the baseline models.

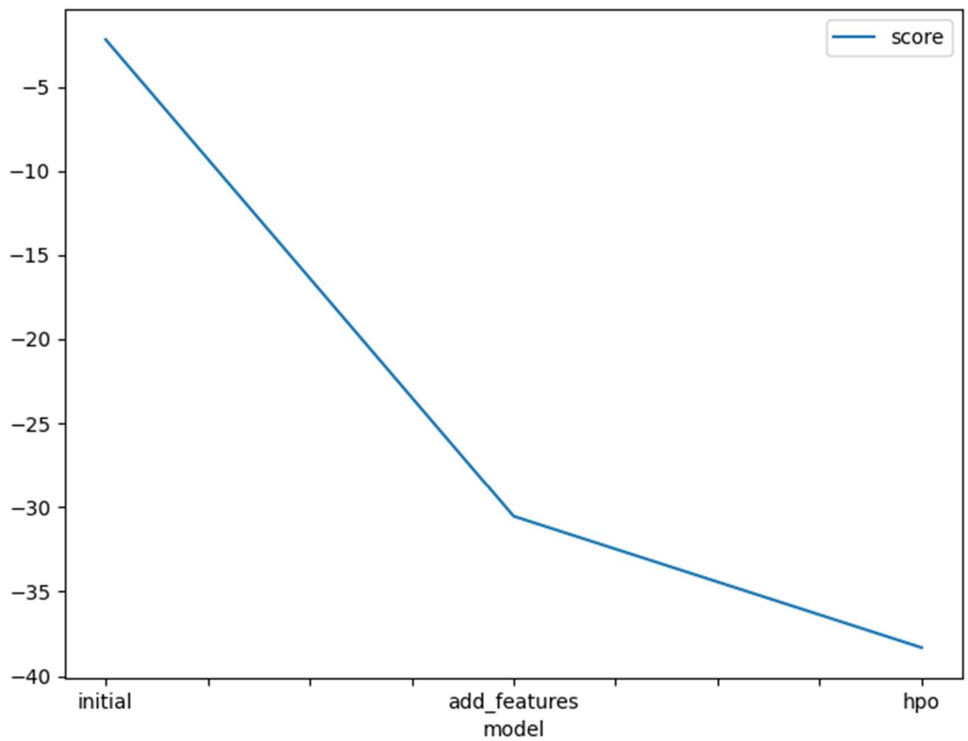
2. If you were given more time with this dataset, where do you think you would spend more time?

TODO: I would spend time with trying different models to train also changing the parameters of the individual models to see the effect on the performance

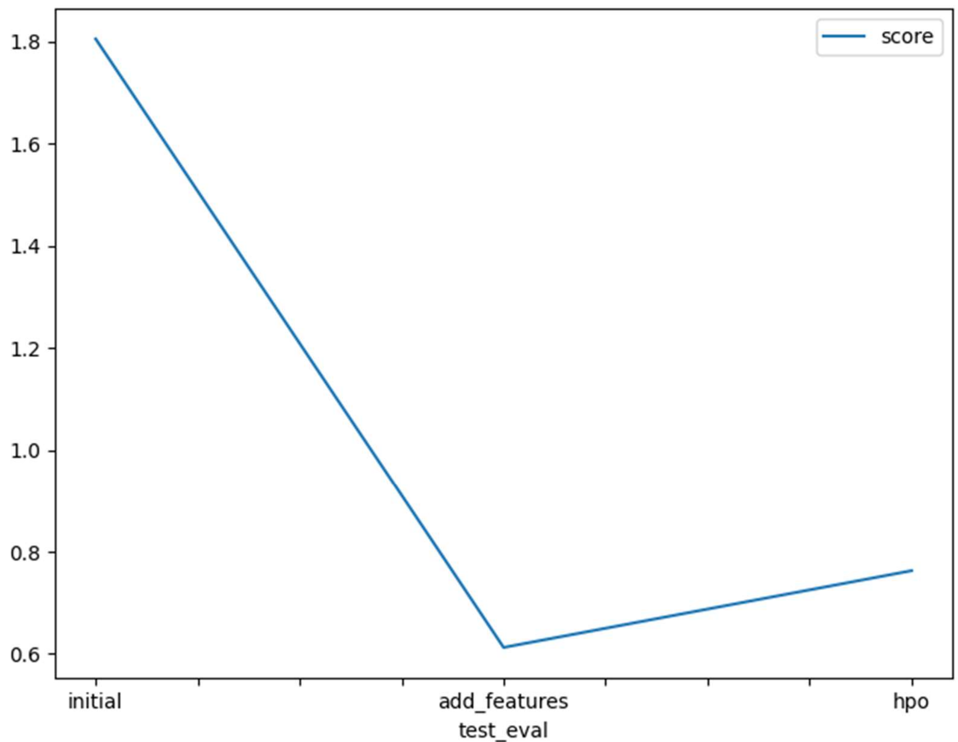
3.Create a table with the models you ran, the hyperparameters modified, and the kaggle score.

	model	hpo1	hpo2	hpo3	score
0	initial	time_limit	presets	None	1.80546
1	add_features	time_limit	presets	None	0.61256
2	hpo	Model hyperparameter	light	time_limit	0.76276

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



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



Summary

TODO: Since the performance or evaluation metric was RMSE so,from the graph for the training runs and also from the performance it is clear that making data for clean and separated the model performance improves a lot also optimising the hyperparameters on the well defined and structured data we can improve the performance of the models further. Based on the performance too the ranking in Kaggle will get better and the less the public score that we get the better our ranking will become.