

```
# Report: Predict Bike Sharing Demand with AutoGluon Solution
#### Lamiaa Hassan Mohammed Zain
```

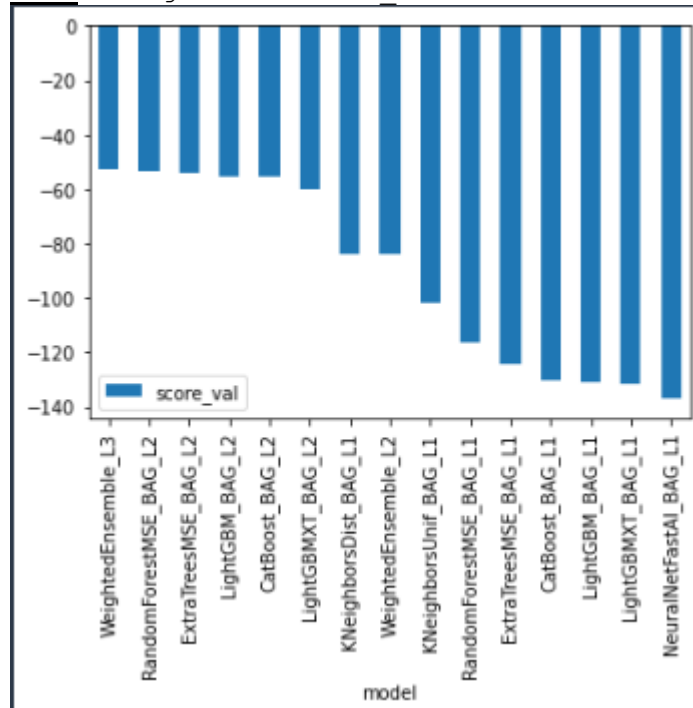
```
## Initial Training
```

```
### 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?
```

```
TODO: Predictions needed to be greater than or equal to 0. That's why
clipping the predictions to have the lower boundary of zero was done.
Also rounding the target 'count'
```

```
### What was the top ranked model that performed?
```

TODO: WeightedEnsemble_L3



```
## Exploratory data analysis and feature creation
```

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

TODO:

- It was found that people rented bikes more from Jan to Feb 2011 and from Sep to Oct 2012. People tend to rent more during the following hours (1-2.5 am, 7-9 am, 1-4 pm, 9-11 pm).
- There was a strong correlation between Season and month, hence the season column was dropped from the training and testing datasets.
- Features like the weather were categorized.

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

TODO: From public score of 1.8, it was improved to be 0.57651.

- This was mainly because of the drop of the correlated column and
- because of the better representation of the datetime column where new features of hour, month and day were used and datetime column was dropped.

```
## Hyper parameter tuning
### How much better did your model preform after trying different hyper
parameters?
```

TODO: It was slightly improved from 0.57651 to 0.55532 as appeared in Kaggle public score.

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

TODO: I would do data normalization and more feature engineering with EDA.

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

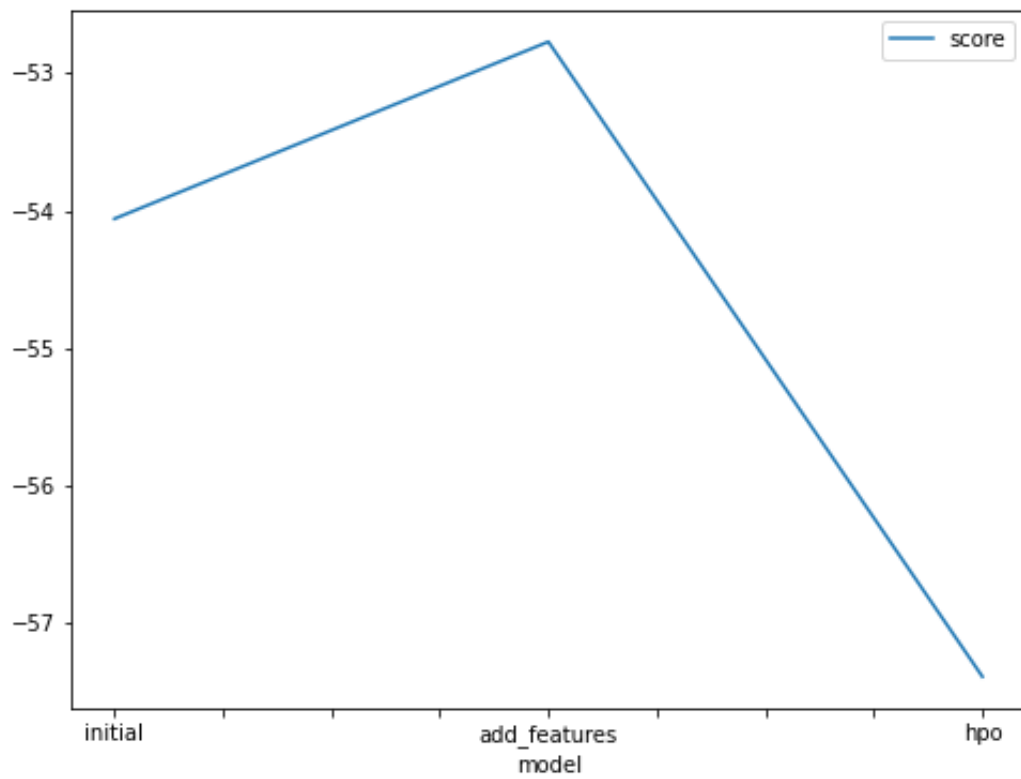
```
|model|hpo1|hpo2|hpo3|score|
|--|--|--|--|--|
|initial|?|?|?|?|
|add_features|?|?|?|?|
|hpo|?|?|?|?|
```

TODO:

model	Hypo1	Hypo2	Hypo3	Hypo4	score
initial	default	default	default		1.81023
add_features	default	default			0.57651
hpo	"'GBM':{'num_boost_round':100,'num_leaves':ag.space.Int(lower=26, upper=66, default=36),'objective':'huber','n_estimators':200}"	'KNN':{'n_neighbors':4,'weights':'distance','algorithm':'brute'}	'XGB': {'objective':'reg:pseudohubererror','eval_metric':'rmse' }	'RF': {'n_estimators':200,'criterion':['gini','entropy','log_loss']}	0.55532

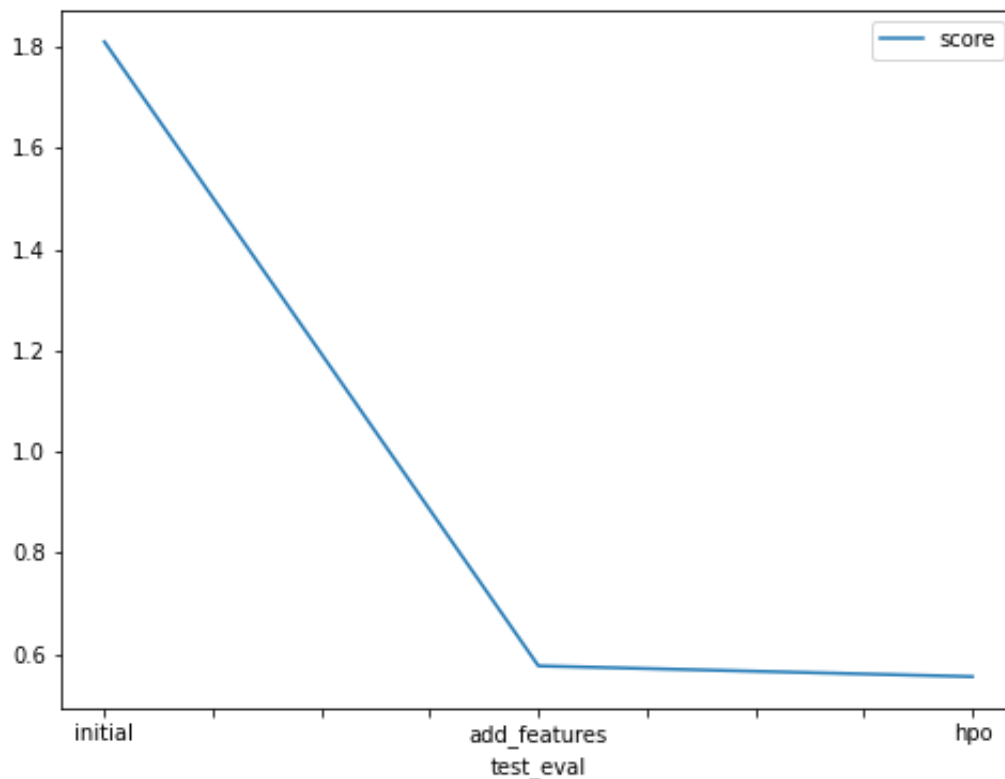
```
### Create a line plot showing the top model score for the three (or
more) training runs during the project.
```

TODO: Replace the image below with your own.



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

TODO: Replace the image below with your own.



Summary

TODO:

In this project, we have obtained/explored following steps in the ML Lifecycle.

- Problem: Better utilization of the data generated by the systems in the bike sharing industry as well as exogenous information are keys to take business advantages for firms.
- Business Objective: Predict bike sharing demand provided data related to different factors provided.
- Obtain Data Data is obtained from the "Bike Sharing Demand" kaggle competition.
- Analyze Data Data is described, visualized and modified mainly using Pandas, Matplotlib and Seabon library.
- Build Model Built models using Autogluon AutomML library. Initially created model without any preprocessing. Later, built models creating new features and tuning hyper parameters.

- Test Model Models are tested by submitting the predictions to Kaggle and compared the models based on the score obtained.