Predicting Bike Rental Counts in Seoul Bike Sharing System

Understanding the Demand for Rental Bikes

Introduction

- The emergence of rental bikes has transformed urban mobility, offering a convenient and sustainable mode of transportation.
- As rental bike usage grows, accurately predicting demand becomes paramount for optimizing the availability and accessibility of bikes.
- Introduction to the dataset: "Today, we delve into the Seoul Bike Sharing System dataset, a treasure trove of information encompassing bike rental counts, weather data, and holiday information."

Problem Statement

The challenge lies in predicting bike counts hour by hour to ensure a stable supply of rental bikes.

Emphasis on the impact: Accurate predictions reduce waiting times, enhancing overall mobility comfort for users.

Dataset Overview

MULTIVARIATE NATURE WITH 14 FEATURES.

BUSINESS FOCUS WITH A REGRESSION TASK.

8760 INSTANCES CAPTURING HOURLY BIKE RENTAL COUNTS.

Motivation

WHY PREDICT BIKE COUNTS?

MOTIVATION LIES IN ENHANCING URBAN MOBILITY

MINIMIZING WAITING TIMES.

ENSURING A STABLE SUPPLY OF RENTAL BIKES

BENEFITS EXTEND TO EFFICIENT SUPPLY CHAIN MANAGEMENT FOR RENTAL BIKES

Variables

- Weather-related data (Temperature, Humidity, Windspeed, etc.).
- Date and time information.
- How weather conditions influence bike rental demand
- The role of date and time in revealing rental patterns

	Date	Rented Bike Count	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)	Seasons	Holiday	Functioning Day
0	01/12/2017	254	0	-5.2	37	2.2	2000	-17.6	0.00	0.0	0.0	Winter	No Holiday	Yes
1	01/12/2017	204	1	-5.5	38	0.8	2000	-17.6	0.00	0.0	0.0	Winter	No Holiday	Yes
2	01/12/2017	173	2	-6.0	39	1.0	2000	-17.7	0.00	0.0	0.0	Winter	No Holiday	Yes
3	01/12/2017	107	3	-6.2	40	0.9	2000	-17.6	0.00	0.0	0.0	Winter	No Holiday	Yes
4	01/12/2017	78	4	-6.0	36	2.3	2000	-18.6	0.00	0.0	0.0	Winter	No Holiday	Yes
5	01/12/2017	100	5	-6.4	37	1.5	2000	-18.7	0.00	0.0	0.0	Winter	No Holiday	Yes
6	01/12/2017	181	6	-6.6	35	1.3	2000	-19.5	0.00	0.0	0.0	Winter	No Holiday	Yes
7	01/12/2017	460	7	-7.4	38	0.9	2000	-19.3	0.00	0.0	0.0	Winter	No Holiday	Yes
8	01/12/2017	930	8	-7.6	37	1.1	2000	-19.8	0.01	0.0	0.0	Winter	No Holiday	Yes
9	01/12/2017	490	9	-6.5	27	0.5	1928	-22.4	0.23	0.0	0.0	Winter	No Holiday	Yes

Methodology

A brief overview of the methodology employed for regression analysis :

Data preprocessing

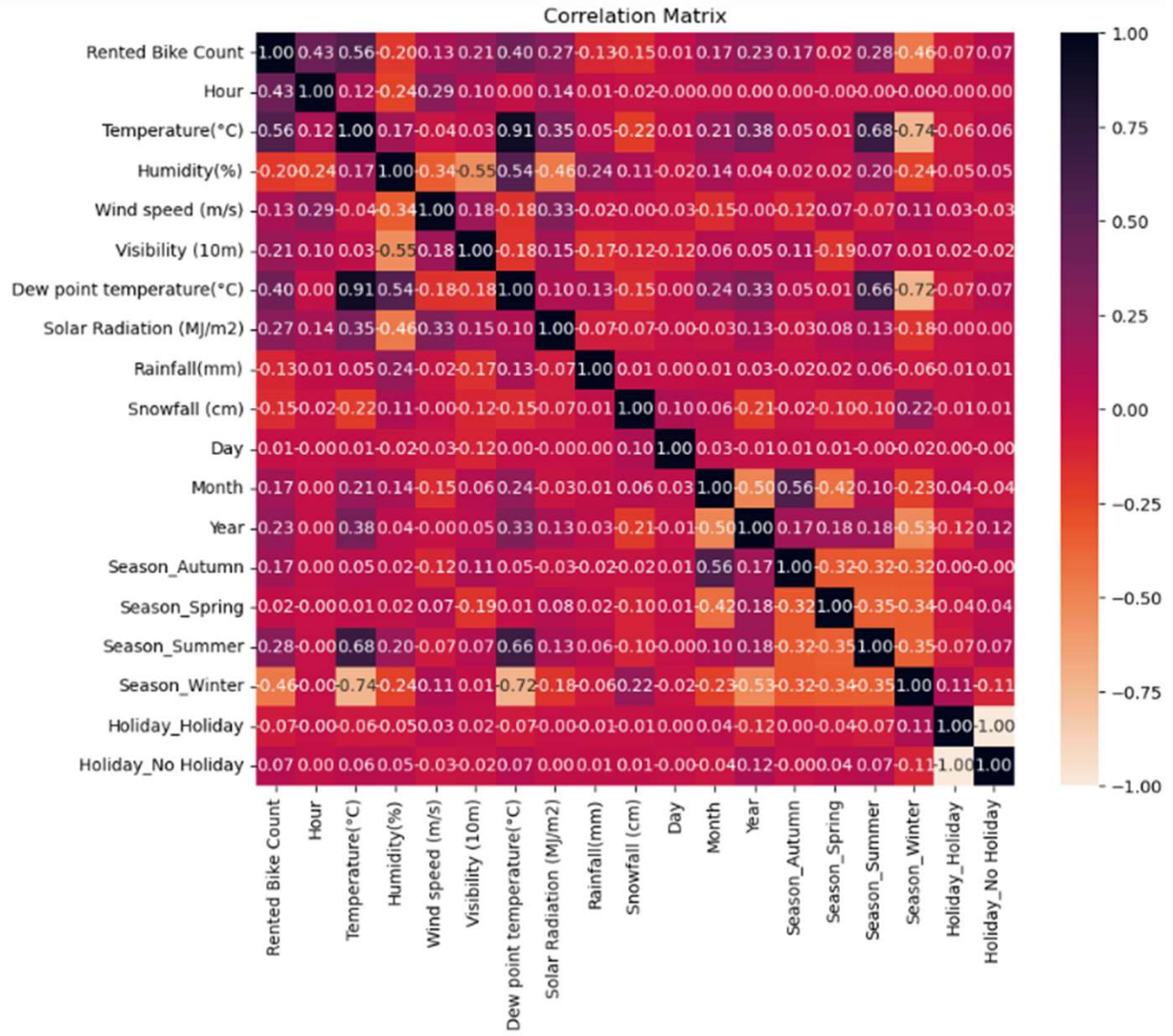
Feature selection

Model training and evaluation

Data Pre-processing

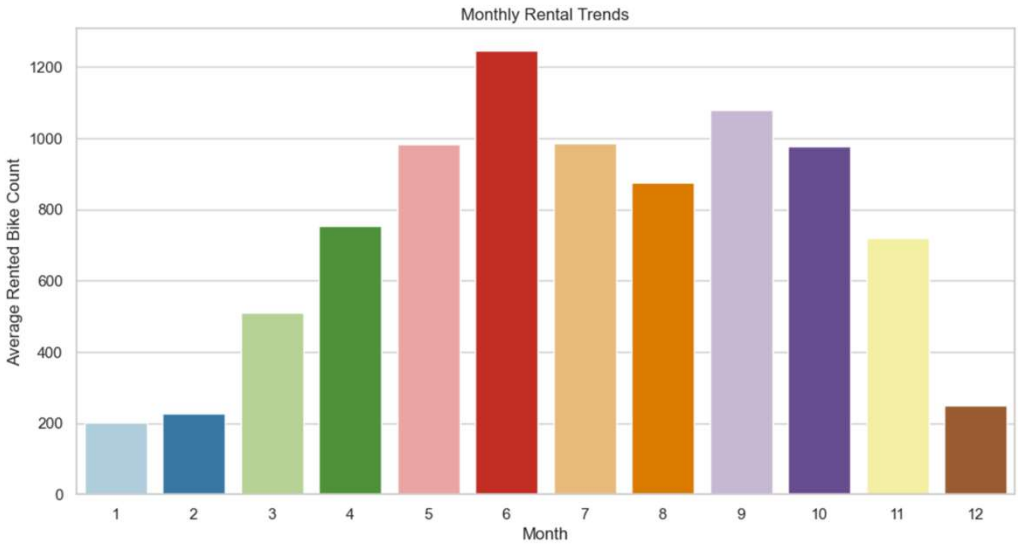
Encoding : Convert categorical features to numerical (seasons, holidays…)

Normalization : Put the numerical variable on the same scale

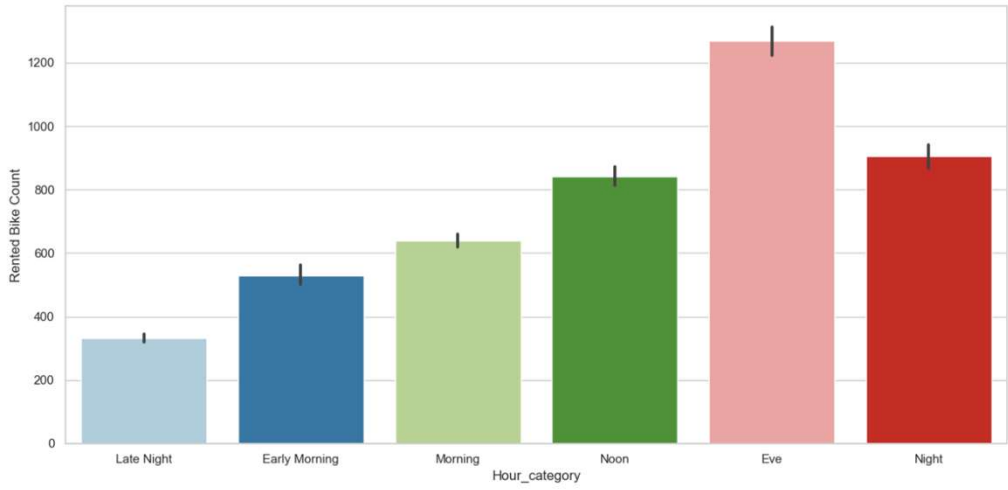


Data Visualization

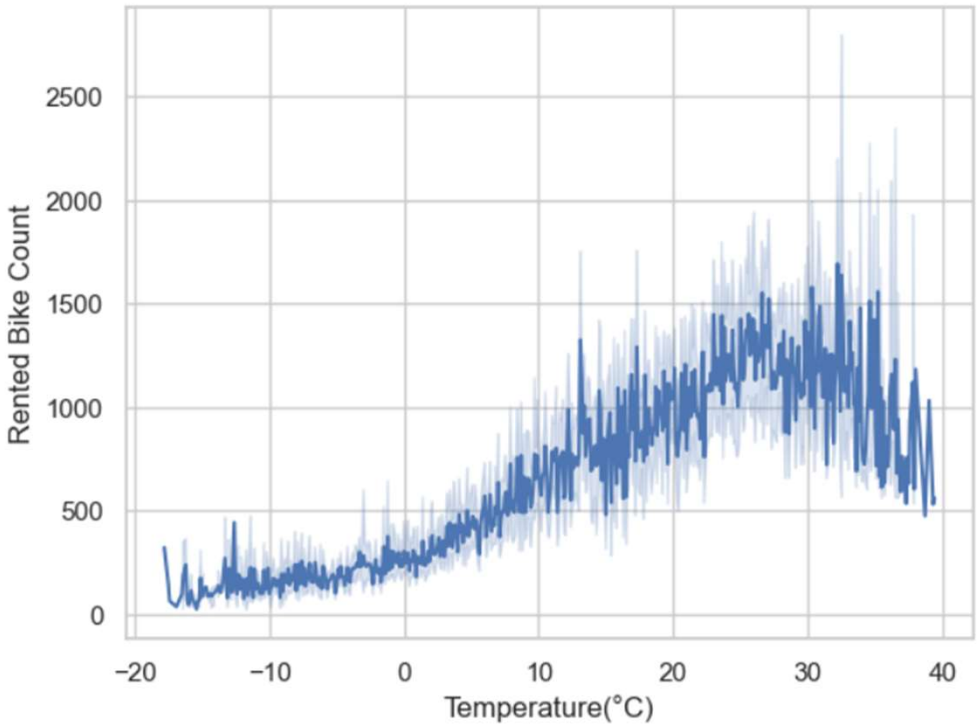
- Average rented bike count per month

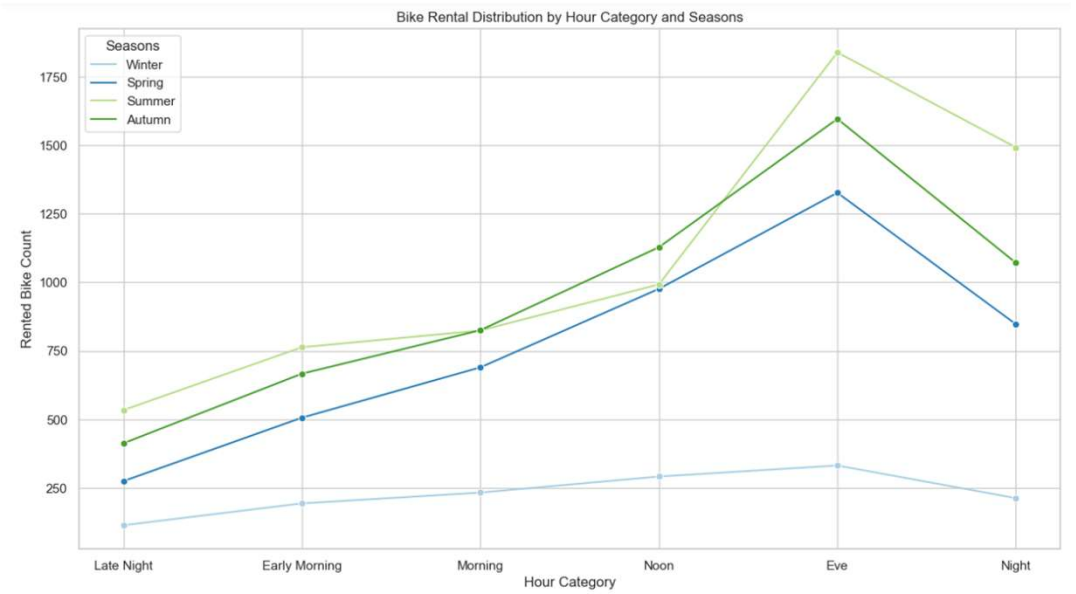


- Average rented bike count per hour category



Data Visualization





Modeling

Implement

Find the best model

Improve

Tune our model

Find the best hyperparameters

Compare

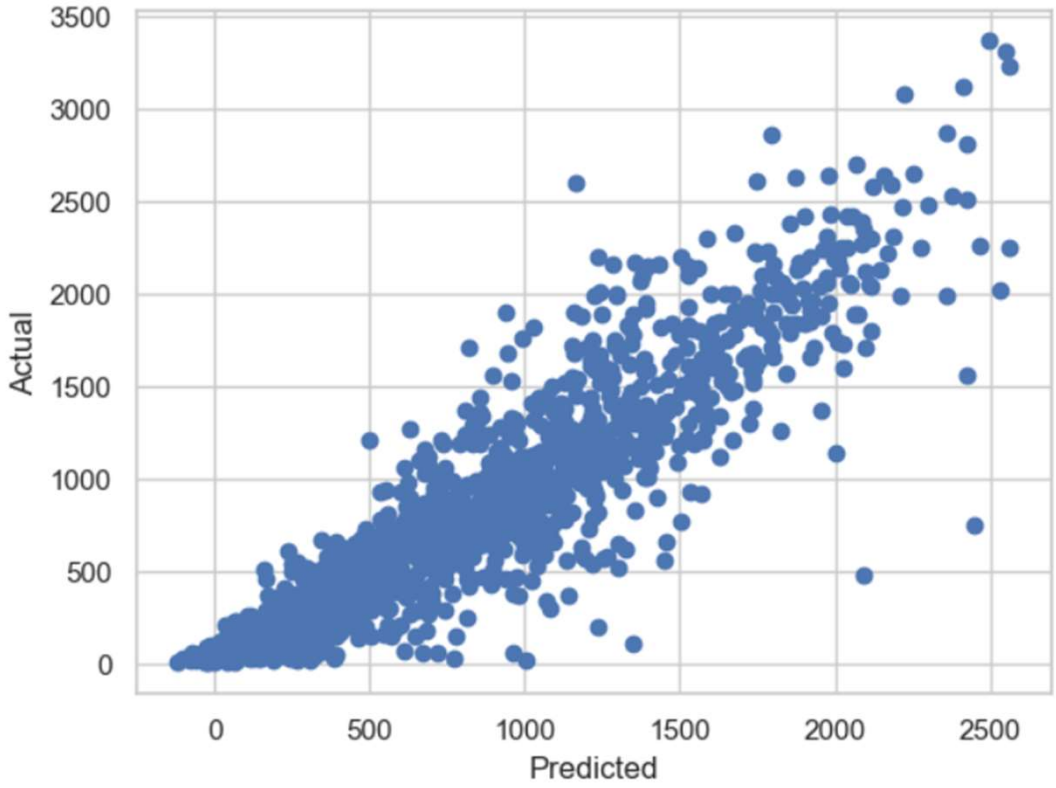
Compare our model before and after tuning

Implement

- Best accuracy (R^2) :

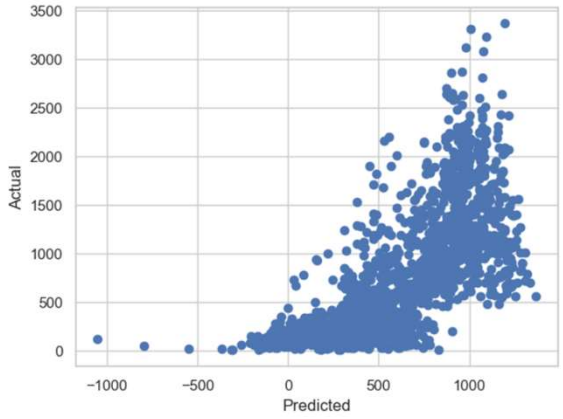
min-max, model:
GradientBoostingRegressor

```
Model Accuracy: 0.851
The mean squared error (MSE) on test set: 58578.6911
Root Mean Squared Error is 242.0304
```

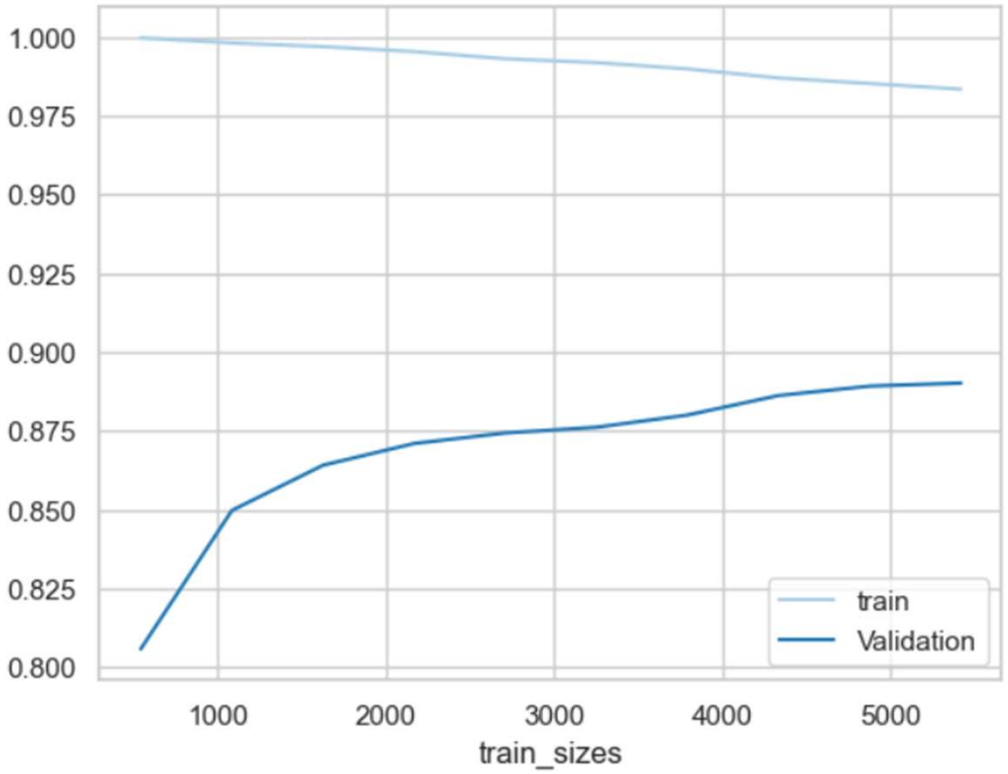


- Exemple of other model:

```
Scaler: RobustScaler, model: LinearSVR
R²=0.4553906860016632
The mean squared error (MSE) on test set: 462.3918
Root Mean Squared Error is 462.3918
```



```
[ 541 1083 1625 2166 2708 3250 3791 4333 4875 5417]
Maximum value : 0.8901894613088303
Reach when n_estimator = 5417
```



Test

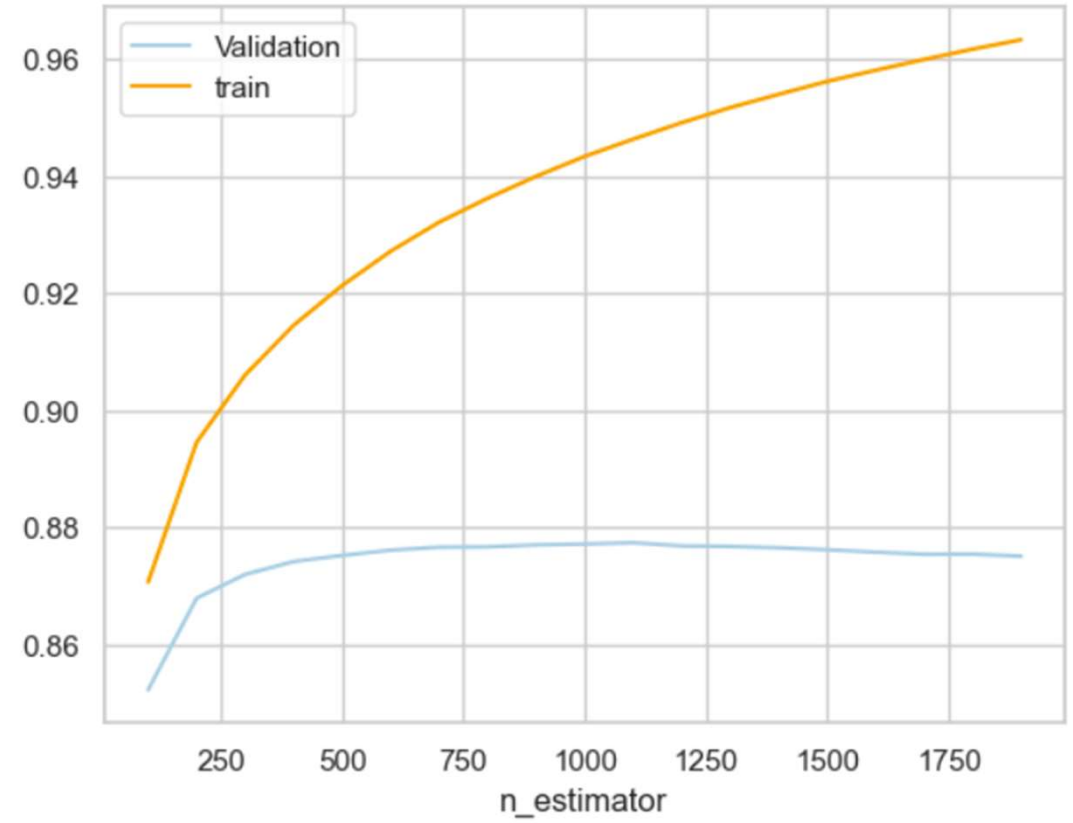
- We test different hyperparameters
- Here we tested the best test size
- test_size= 0.36 (= 1- (5417/8465))

Validation Curve

Improve model one parameter at a time

Not optimal

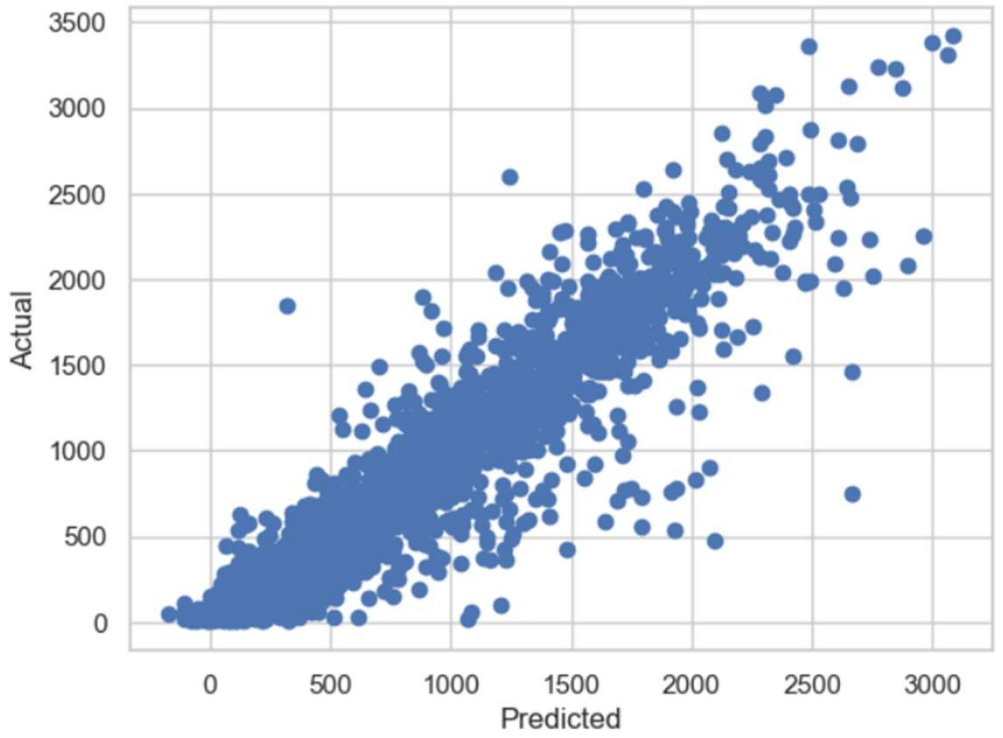
```
Maximum value : 0.877395919653949
Reach when n_estimator = 1100
```



Improve / Grid Search

Using hyperperameters tuning (Grid Search)

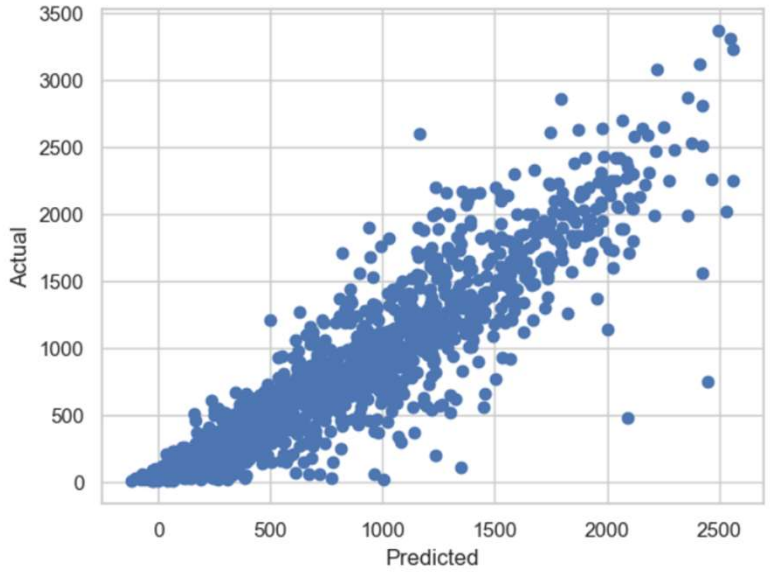
```
Model Accuracy: 0.891
The mean squared error (MSE) on test set: 44343.0550
Root Mean Squared Error is 210.5779
```



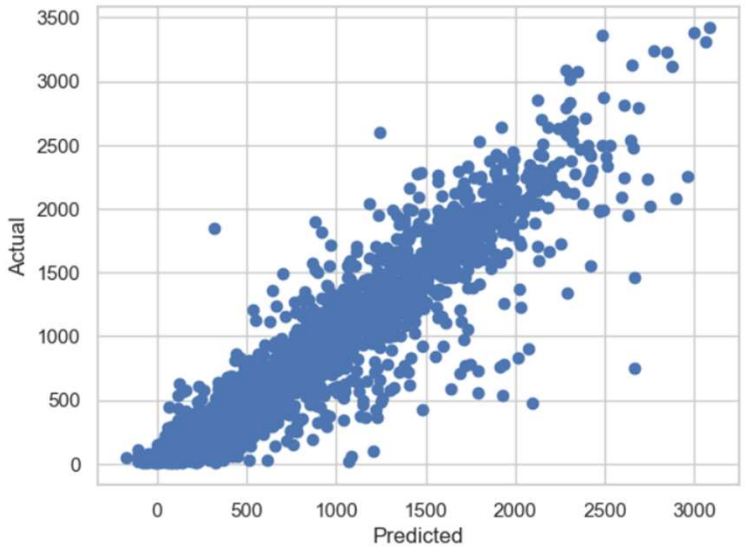
Results

we improved accuracy of the model from 85% to 89%

```
Model Accuracy: 0.851
The mean squared error (MSE) on test set: 58578.6911
Root Mean Squared Error is 242.0304
```




```
Model Accuracy: 0.891
The mean squared error (MSE) on test set: 44343.0550
Root Mean Squared Error is 210.5779
```



Challenges and Limitations

- Model limitations : high margin of error
- Discussion of the study's limitations : only for Korea bike rentals
- To go further : generate predictions