7 Days Bike Usage Forecasting for Bike-Sharing System Management

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Lin Yu-Chun
Advised by Prof. Qifeng Chen
Supervised by Mouse Ho, Sengital Limited

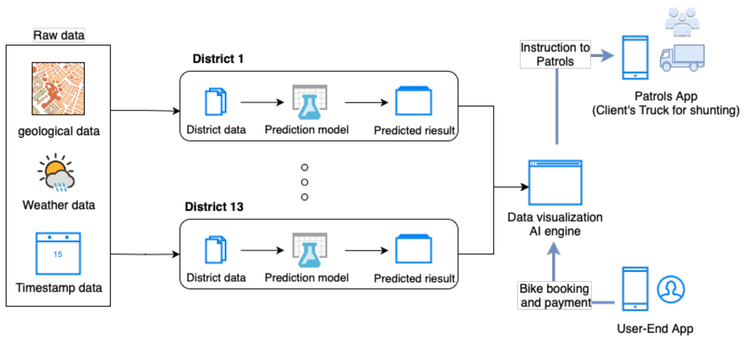
Overview

With the increment in the sharing bike demand, resource distribution and bike management becomes a crucial step among the logistics in bike-sharing system.
However, relying on a 1-day prediction is insufficient to handle the large volume of user activities. By implementing a 7-day bike usage forecast, the patrol team can effectively plan ahead for future conditions.

Objectives

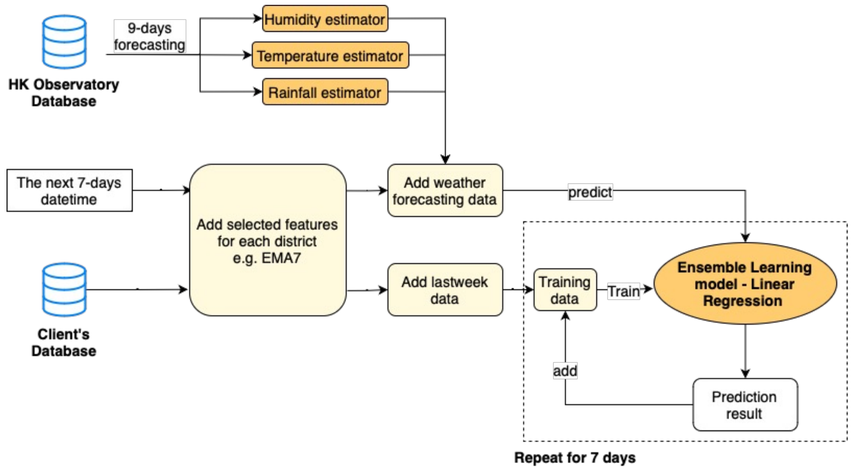
- ✓ Create a reproducible trained model pipeline.
- ✓ Select optimal model for each district.
- ✓ Explore valuable feature to leverage data for each district.
- ✓ Raise the accuracy of model compared to 1-day prediction.

System Design



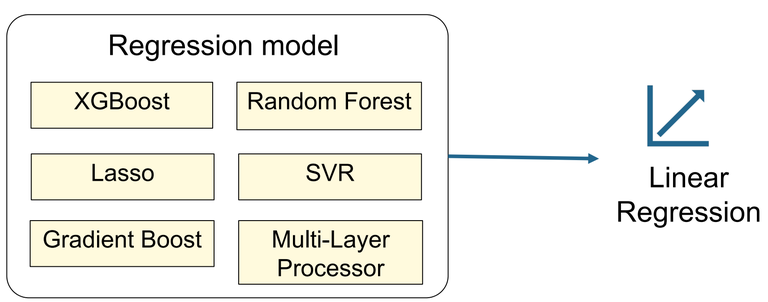
1. Process the raw data
2. Feature engineering
3. Recursive multi-step forecasting
4. Result uploads to the client's database
5. Result becomes accessible to administrative and patrol team members.

Implementation



We obtained the upcoming 9 days weather forecast report from the HK Observatory and the client's historical data to train the model. We then incorporated selected features for each district based on SHAP analysis. Finally, we trained a linear regression model using the selected base models, which were chosen independently based on their performance scores.

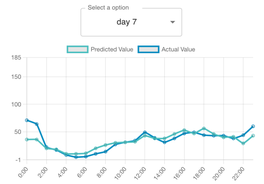
How Does Our Model Work?



We tested **6 regression models** and conducted experiments for each district's data to identify the optimal combination. The base model initially makes predictions, and then the results are connected to Linear Regression to produce the final result.

Evaluation

The example result shows the prediction for Tai Po from 2023-10-01 to 2023-10-07.



Based on results, we can conclude that our system has successfully provided accurate forecasts for the next 7-day bike usage using the ensemble learning method.

Best Model for each district

District no.	model 1	model 2	model 3	R-Squared	RMSE
1	svr	mlp	lasso	0.8574	19.1814
2	svr	mlp	lasso	0.0345	0.8938
3	svr	gbr	lasso	0.7547	4.1960
4	svr	xgboost	lasso	0.6960	4.4612
5	svr	rfr	lasso	0.7691	2.2581
6	svr	mlp	lasso	0.8044	10.8145
7	svr	gbr	xgboost	0.3621	0.9040
8	svr	gbr	lasso	0.8161	8.1783
9	svr	gbr	lasso	0.6109	3.9703
10	rfr	xgboost	mlp	0.6914	5.7696
11	svr	gbr	xgboost	0.4780	1.5407
12	svr	gbr	xgboost	0.5868	3.1304
13	rfr	xgboost	mlp	0.8089	3.7868