

FLIP₀₀ FINAL PRESENTATION

Zhaoyang Wang
Xi'an Shiyou University, China

Introduction

You are given 5 years of store-item sales data, and asked to predict 3 months of sales for 50 different items at 10 different stores.

date - Date of the sale data. There are no holiday effects or store closures.

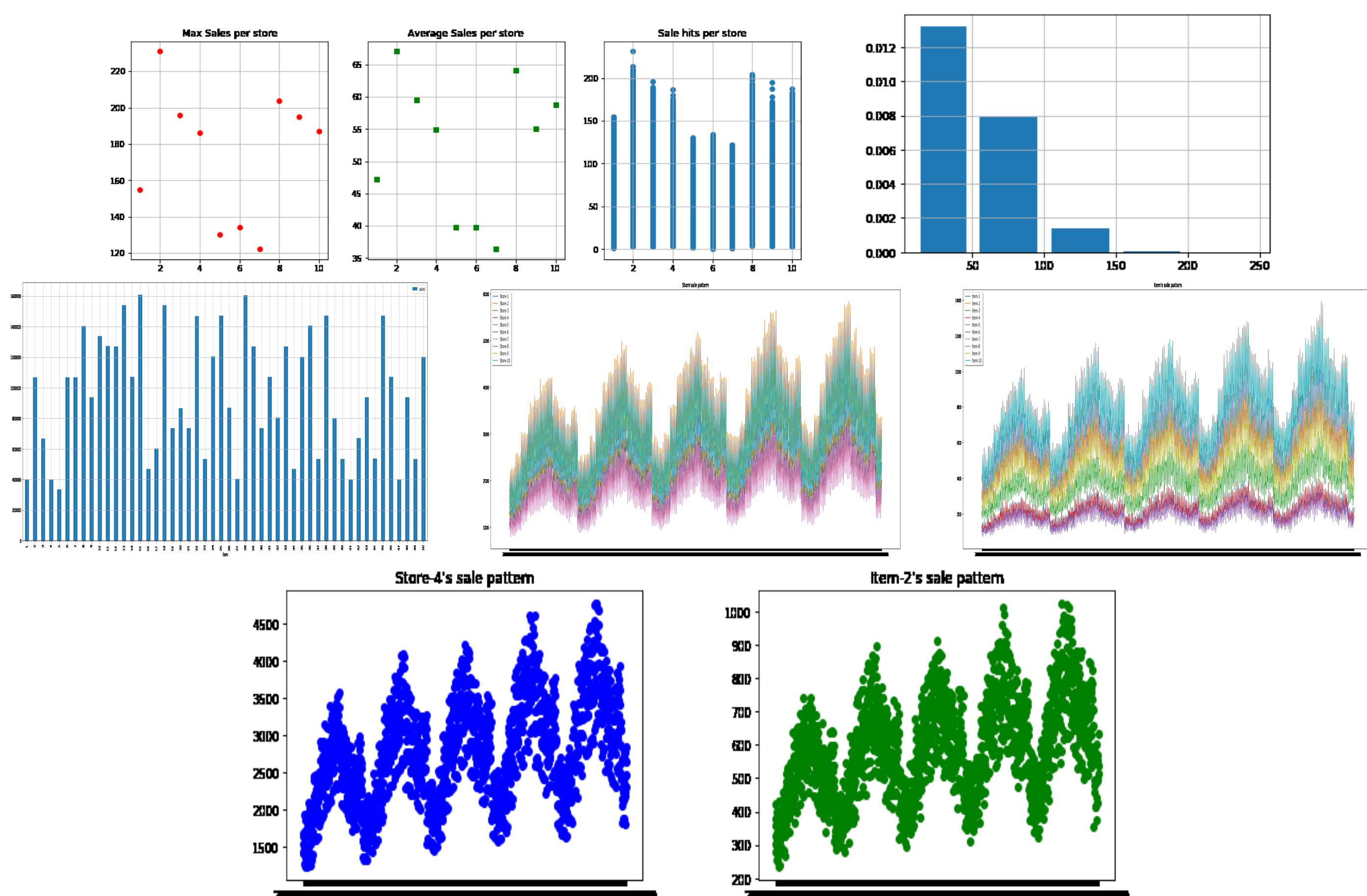
store - Store ID

item - Item ID

sales - Number of items sold at a particular store on a particular date.

Data Visualization

By using matplotlib to describe the data. And Observe the relationship between different features.



Data Preparation

Through the data visualization before, we can intuitively recognize the changes in sales. However, to forecast sales for the next three months, we need to extract some new features. From the previous figure we can see that the sales are related to the characteristics of the year, month, season, etc., so we can add some new features.

dayofweek - The day of the week. Monday is indicated by 0, Tuesday is indicated by 1, and so on.

is_weekend - Determine if this day is a weekend.

day - The day of the month.

year - Judging year.

dayofyear - The day of the year.

weekofyear - The week of the year.

sales_mean_lag_90 - Calculate 90 days from the day before, and then start from this day, the average of the first seven days.

sales_std_lag_90 - Calculate 90 days from the day, and then start from this day, the standard deviation of the first seven days.

Algorithm

There are many machine learning methods for solving regression problems. This moment I will chose the lightGBM model. The lightGBM is light Gradient Boosting Machine which is a gradient boosting framework that uses tree based learning algorithms.

Faster training speed and higher efficiency.

Lower memory usage.

Better accuracy.

Support of parallel and GPU learning.

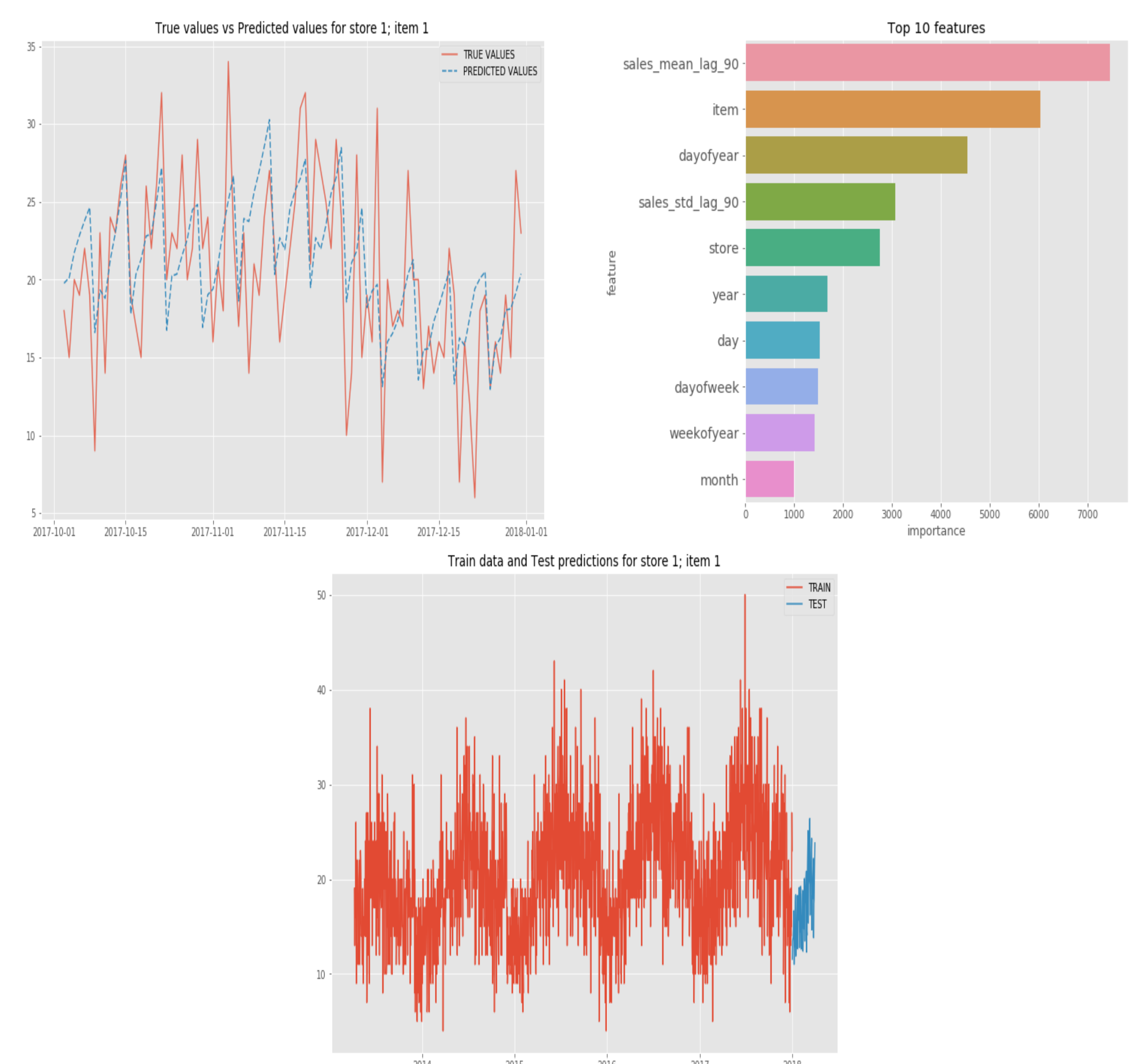
Capable of handling large-scale data

Choosing the SMAPE as Evaluate model

$$SMAPE = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

Forecasting result

Figure of left is the first prediction based on the model. Figure of center shows ten features with high feature importance. Figure of the right is a new prediction based on the model to get the results needed for this problem.



Conclusion

Data Analysis Using visual methods to find connections and features within the dataset.

Feature Engineering Find and extract important features.

Modeling Choose the suitable model parameters.

Prospecting I would like to select multiple models.

Acknowledgement
• Thank you