Flipoo Project Report

Yao Yang

Chongqing University of Posts and Telecommunications

Introduction

The goal of this project is to forecast bike rental demand given the input feature like *weather*, *temperature*, *humidity*, and *windspeed*.
The raw datasets contain eight data fields, which attributes are shown below.

Name	Attribute
datetime	hourly date + timestamp
season	1 = spring, 2 = summer, 3 = fall, 4 = winter
holiday	whether the day is considered a holiday
workingday	whether the day is neither a weekend nor holiday
weather	1: Clear, 2: Mist + Cloudy, 3: Light Snow, 4: Heavy Rain
temp	temperature in Celsius
atemp	"feels like" temperature in Celsius
humidity	relative humidity
windspeed	wind speed
casual	number of non-registered user rentals initiated
registered	number of registered user rentals initiated
count	number of total rentals

Data Preprocessing and Feature EnginPeering

- Process date data (datetime module)
- Transform categorical features (calendar module)
- Analyze missing value and handle outlier
- Analyze target variable (logarithmic transformation)
- Fill in zero values in the windspeed feature (random forest model)

Build and Solve the Model

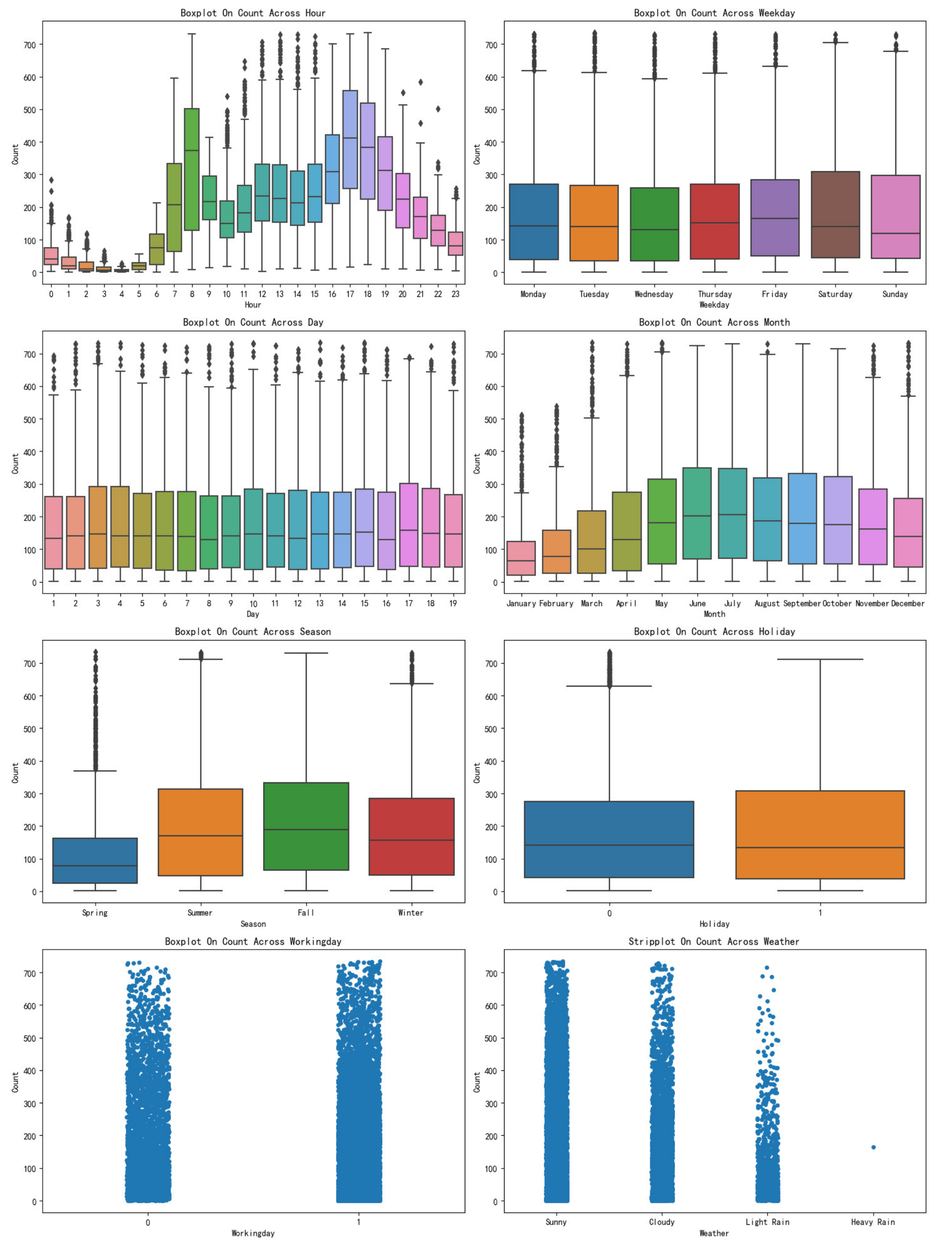
	Model	RMSLE
15	LightGBM	0.316161
11	RandomForestRegressor	0.375379
10	BaggingRegressor	0.394187
14	XGBoost	0.422559
13	GBRT	0.435759
8	DecisionTreeRegressor	0.523695
9	ExtraTreeRegressor	0.554145
12	AdaBoostRegressor	0.697286
4	KernelRidge Regression	0.813210
7	KNN	0.864965
6	SVR	1.045943
5	ElasticNet Regression	1.053736
3	Ridge Regression	1.053749
2	Lasso Regression	1.054156
0	Linear Regression	1.054414
1	Logistic Regression	1.127804

Modeling and Result

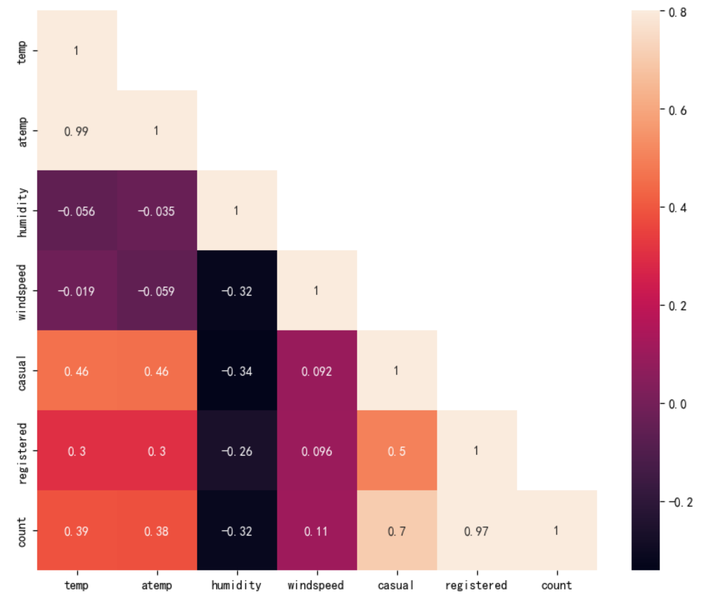
- Model: Stacking*0.6+LightGBM*0.4
- Score: 0.50413
- Rank: 1741/3243

Data Visualization

The eight category features are visualized.



Plot a heat map of the correlation matrix between the individual numerical features.



Conclusion

When modeling, we mainly consider the three numerical features "temp", "humidity" and "windspeed". Sixteen basic machine learning regression prediction family models were used. The final prediction is composed of the predictions of the best two models Stacking and LightGBM with a weight of 0.6 and 0.4.