BIKE SHARING DEMAND FORECAST USE OF A CITY BIKESHARE SYSTEM

MIRIAM ZHU

Abstract. Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city. Using these systems, people are able rent a bike from a one location and return it to a different place on an as-needed basis. Currently, there are over 500 bike-sharing programs around the world.

The data generated by these systems makes them attractive for researchers because the duration of travel, departure location, arrival location, and time elapsed is explicitly recorded. Bike sharing systems therefore function as a sensor network, which can be used for studying mobility in a city. In this competition, participants are asked to combine historical usage patterns with weather data in order to forecast bike rental demand in the Capital Bikeshare program in Washington, D.C.

Contents

Date: (None).
2020 *Mathematics Subject Classification.* Artificial Intelligence.
Key words and phrases. Machine Learning, Bike sharing, Kaggle.

1. Introduction

Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city. Using these systems, people are able rent a bike from a one location and return it to a different place on an as-needed basis. Currently, there are over 500 bike-sharing programs around the world.

The data generated by these systems makes them attractive for researchers because the duration of travel, departure location, arrival location, and time elapsed is explicitly recorded. Bike sharing systems therefore function as a sensor network, which can be used for studying mobility in a city. In this competition, participants are asked to combine historical usage patterns with weather data in order to forecast bike rental demand in the Capital Bikeshare program in Washington, D.C.

2. Dataset Description

- The competition provides hourly rental data spanning two years. For this competition, the training set is comprised of the first 19 days of each month, while the test set is the 20th to the end of the month. You must predict the total count of bikes rented during each hour covered by the test set, using only information available prior to the rental period.
- Data Fields

Attribute	Description
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, 3: Light Snow, Light Rain , 4: Extreme weather
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

3. Data Visualization

- Time characteristic analysis

4. Feature Selection and Feature Importance

- The influence of characteristics on count is as follows: hour¿temp¿atemp¿humidity¿month¿season¿year¿weather ¿windspeed¿workingday¿weekday¿day¿holiday
- Weather characteristics analysis

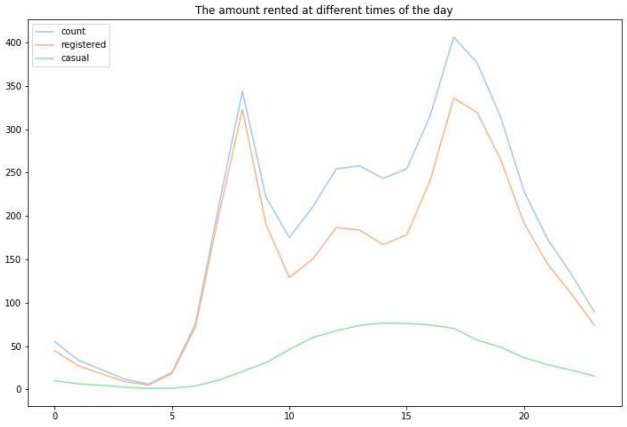

FIGURE 1. The amount rented at different times of the day

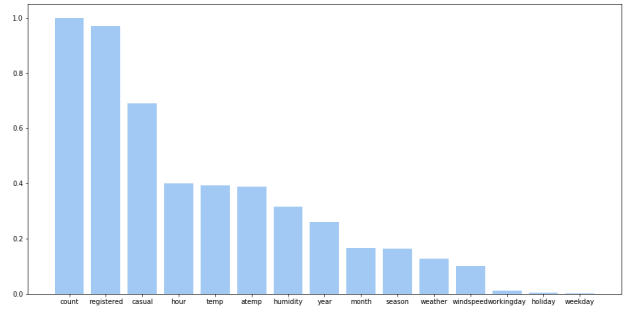

FIGURE 2. Correlation rank

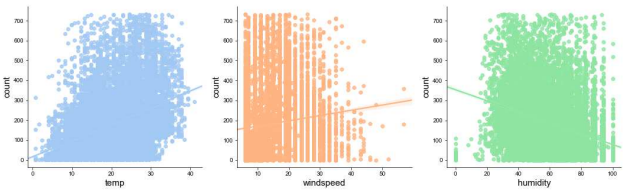

FIGURE 3. The effect of weather on rental amount

5. MODELING AND RESULT

Model: RandomForestRegressor

Result	The initial model	After cross validation
Accuracy	0.9338	0.9249
MSLE	0.0152	0.0159

6. Conclusion

- Cross-validation by grid search did not decrease the RMSE of the model and did not improve the accuracy of the model. The effect did not come up to expectations.
- The limitation of this study is that it does not consider whether there is overfitting of the model, and further experiments can be carried out in future studies.

(A. 1) School of Data Science,, Deakin University, Burwood 3125, Australia
Email address, A.: zhudo@tulip.academy