

# Final Report

Group 1: Jiawen Chen, Brooke Felsheim, Elena Kharitonova, Jairui Tang, and Xinjie Qian

4/29/2022

## Introduction

```
library(usethis)
library(devtools)
load_all("package/bikeSharing/")

## i Loading bikeSharing

set.seed(735)

str(seoul)

## 'data.frame':    1059 obs. of  13 variables:
## $ Date           : chr  "01-01" "01-01" "01-01" "01-02" ...
## $ Hour_chunks    : Factor w/ 3 levels "[0,8)","[8,16)",...: 1 2 3 1 2 3 1 2 3 1 ...
## $ Day            : num   1 1 1 2 2 2 3 3 3 4 ...
## $ Is_weekend      : num   0 0 0 0 0 0 0 0 0 0 ...
## $ Is_holiday      : num   1 1 1 0 0 0 0 0 0 0 ...
## $ Season          : chr   "Winter" "Winter" "Winter" "Winter" ...
## $ Min_temp        : num  -5 -5 -5 -3.8 -3.8 -3.8 -7 -7 -7 -8.6 ...
## $ Max_temp        : num   3.7 3.7 3.7 1.7 1.7 1.7 -0.4 -0.4 -0.4 -0.8 ...
## $ Min_humidity    : int   20 20 20 20 20 20 29 29 29 31 ...
## $ Max_humidity    : int   56 56 56 71 71 71 54 54 54 57 ...
## $ Wind_speed      : num   0.9 1.85 1.61 0.65 2.26 ...
## $ Rain_or_snow    : num   0 0 0 0 0 0 0 0 0 0 ...
## $ Bike_count      : int  1002 1633 1655 938 2610 2898 1022 2624 2866 1015 ...

str(london)

## 'data.frame':    2188 obs. of  13 variables:
## $ Date           : chr  "01-01" "01-01" "01-01" "01-01" ...
## $ Hour_chunks    : Factor w/ 3 levels "[0,8)","[8,16)",...: 1 1 2 2 3 3 1 1 2 2 ...
## $ Day            : num   1 1 1 1 1 1 2 2 2 2 ...
## $ Is_weekend      : num   0 1 0 1 0 1 0 1 0 1 ...
## $ Is_holiday      : num   1 0 1 0 1 0 1 0 1 0 ...
## $ Season          : chr   "Winter" "Winter" "Winter" "Winter" ...
## $ Min_temp        : num   3 5 3 5 3 5 1 9 1 9 ...
## $ Max_temp        : num   9 10 9 10 9 10 6 11.5 6 11.5 ...
## $ Min_humidity    : num  76 81 76 81 76 81 71 82 71 82 ...
## $ Max_humidity    : num  87 93 87 93 87 93 93 94 93 94 ...
## $ Wind_speed      : num   2.48 3.65 4.83 4.08 6.63 ...
## $ Rain_or_snow    : num   0 1 1 1 1 1 0 1 0 1 ...
## $ Bike_count      : int  2715 2962 4460 2450 2622 1009 438 475 7756 4263 ...
```

```
str(dc)
```

```
## 'data.frame': 2187 obs. of 13 variables:
## $ Date : chr "01-01" "01-01" "01-01" "01-01" ...
## $ Hour_chunks : Factor w/ 3 levels "[0,8)","[8,16)",...: 1 1 2 2 3 3 1 1 2 2 ...
## $ Day : num 1 1 1 1 1 1 2 2 2 2 ...
## $ Is_weekend : num 1 1 1 1 1 1 0 1 0 1 ...
## $ Is_holiday : int 0 0 0 0 0 0 1 0 1 0 ...
## $ Season : chr "Spring" "Spring" "Spring" "Spring" ...
## $ Min_temp : num 1.4 4.22 1.4 4.22 1.4 4.22 2.34 2.34 2.34 2.34 ...
## $ Max_temp : num 13.6 14.6 13.6 14.6 13.6 ...
## $ Min_humidity: num 72 48 72 48 72 48 32 39 32 39 ...
## $ Max_humidity: num 94 93 94 93 94 93 45 100 45 100 ...
## $ Wind_speed : num 0.208 1.458 3.958 3.75 4.791 ...
## $ Rain_or_snow: num 0 0 0 0 1 1 0 1 0 1 ...
## $ Bike_count : int 108 290 508 1218 369 786 96 55 1102 452 ...
```

## Methods

### Negative Binomial Generalized Linear Mixed Model

### Random Forest

## Results

### Negative Binomial Generalized Linear Mixed Model

### Random Forest

```
set.seed(735)
rf.fit <- train_random_forest(data = seoul)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
rf.fit
```

```
## Random Forest
```

```
##
```

```
## 1059 samples
```

```
## 11 predictor
```

```
##
```

```
## No pre-processing
```

```
## Resampling: Cross-Validated (5 fold)
```

```
## Summary of sample sizes: 847, 847, 847, 848, 847
```

```
## Resampling results across tuning parameters:
```

```
##
```

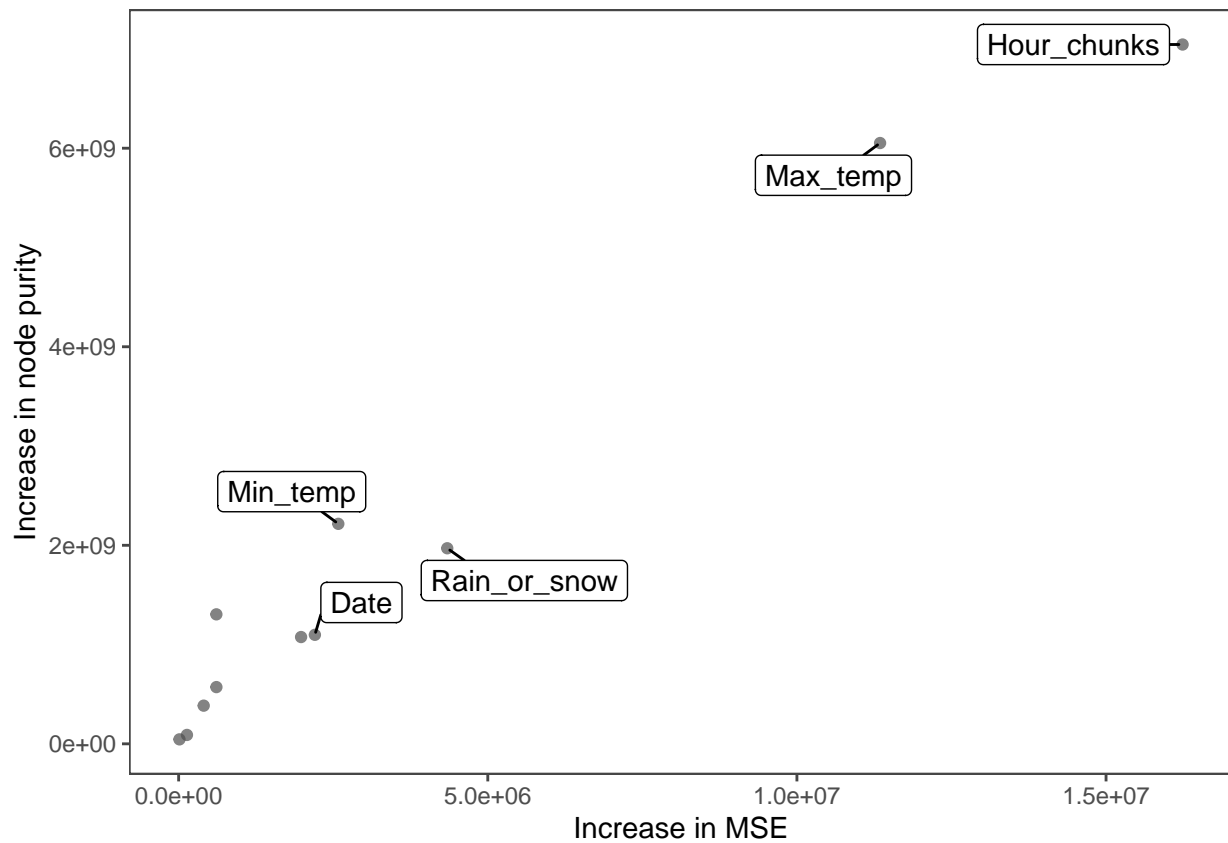
##	mtry	RMSE	Rsquared	MAE
##	2	2015.252	0.8544870	1320.067
##	6	1737.058	0.8778998	1118.894
##	11	1772.282	0.8642370	1142.586

```
##
```

```
## RMSE was used to select the optimal model using the smallest value.
```

```
## The final value used for the model was mtry = 6.
```

```
plot_rf_importance(seoul)
```



```
rf_model_fit(rf.fit, seoul, scale_seoul_mean = "no")
```

```
##      RMSE      MAE      R2
## 1 525.7923 314.1053 0.9873957
```

```
rf_model_fit(rf.fit, london, scale_seoul_mean = "yes")
```

```
##      RMSE      MAE      R2
## 1 3432.079 2738.26 0.6608242
```

```
rf_model_fit(rf.fit, dc, scale_seoul_mean = "yes")
```

```
##      RMSE      MAE      R2
## 1 711.5196 553.0485 0.5998493
```

## Discussion