Predictive Analysis of Bike-share Rental Demand

Benjamin Fillmore

University of Oklahoma

benjaminpaynefillmore@ou.edu

April 29, 2021

Overview

The data for this project came from a Kaggle competition prompting the application of Machine Learning tools to predict bike share rental demand.

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, Few clouds, Partly cloudy, Partly cloudy
2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
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

Bullet Points

- Bike Share programs make bike commuting more accessible
- Remove price barrier to entry
- Dependent on municipal support and weather conditions

Existing Research

Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto

This paper is unique in that it focuses on a relatively succesful bike share program, which function year-round, despite the cold average temperature. This paper builds around socio-demographic influence, as well as the effects of greater incorporation of bike-minded structural changes on active commuting probability.

Bike Share Demand Prediction using RandomForests

This paper focuses on prediction using Random Forests, but equally is primarily concerned with the effects of pollution in urban areas on active commuting probability, using urban centers in India as a benchmark.

Existing Research Cont.

A review on bike-sharing: The factors affecting bike-sharing demand

This paper focuses on the environmental implications of bike share programs, but levies that we will not be able to adopt large-scale bike share programs without understanding the influencing factors.

Community mobility MAUP-ing: A socio-spatial investigation of bikeshare demand in Chicago

This is a unique analysis, because it focuses on an aggregate approach, which emphasizes socio-demographic impact of bike share programs, rather than as a contributing factor. This paper focuses on the measurement of outcome and equity impacts.

Cleaning and Forecasting

Cleaning Process

1. Mutate as.factor
2. Separate Time Elements
3. Distinguish "Workday" from "Weekend"

I used a tuned RandomForest model to make predictions based on all data input, accommodating time, place within week, and multiple weather condition factors.

Cleaner Function

```
tune <- function(df) {
  names = c("season", "holiday", "workingday", "weather")
  df[,names] = lapply(df[,names], factor)
  #
  df$season=as.factor(df$season)
  df$weather=as.factor(df$weather)
  df$holiday=as.factor(df$holiday)
  df$workingday=as.factor(df$workingday)
  df$hour=substr(df$datetime,12,13)
  df$hour=as.factor(df$hour)
  return(df)

}

train    <- tune(train)
test     <- tune(test)
```

This function is essentially applying cleaner names to the columns and then applying the as.factor() command to variables, making them more intuitive for this model.

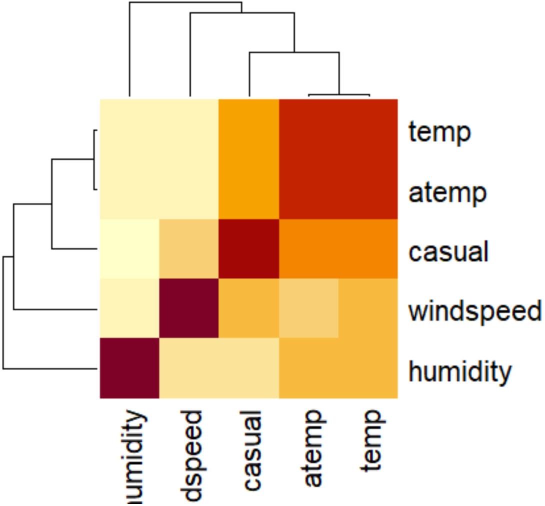
Random Forest Functions

$$
\begin{aligned}
log(reg) = {} & beta_1 datetime + beta_2 season + beta_3 holiday \\
& + beta_4 workingday + beta_5 weather + beta_6 temp \\
& + beta_7 atemp + beta_8 humidity + beta_9 windspeed \\
& + beta_{10} day + beta_{11} year + beta_{12} weekend
\end{aligned}
$$

$$
\begin{aligned}
log(cas) = {} & beta_1 datetime + beta_2 season + beta_3 holiday + \\
& beta_4 workingday + beta_5 weather + beta_6 temp \\
& + beta_7 atemp + beta_8 humidity + beta_9 windspeed + \\
& beta_{10} day + beta_{11} year + beta_{12} weekend
\end{aligned}
$$

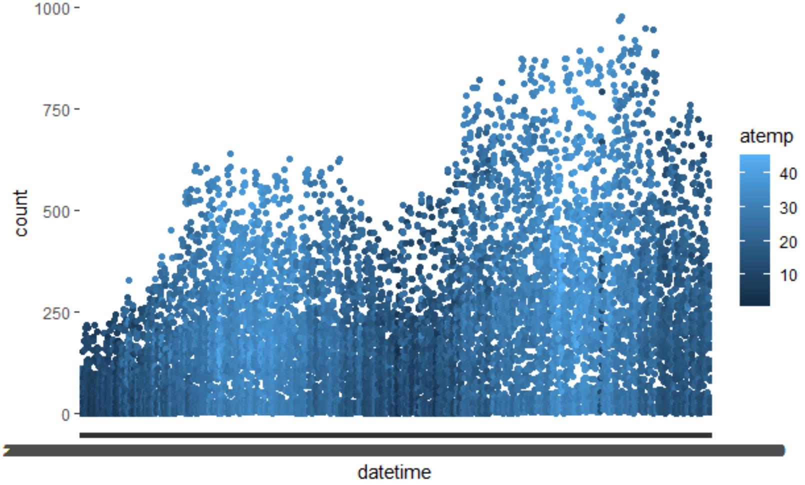
CorMat Heat Map

Shows the relationship between different variables



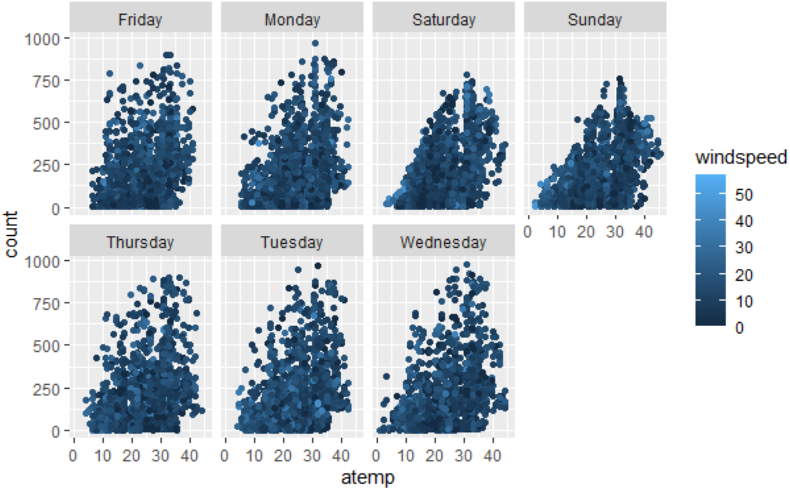
Datetime

Shows the effect of "feels like" weather on rental count, broken down by datetime - which incorporates hour and day



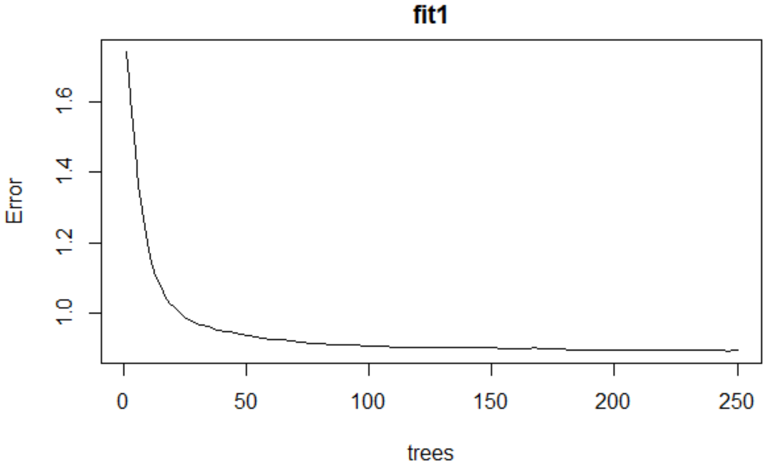
Weekly Demand

Shows the effect of "feels like" weather on rental count, broken down by day of week



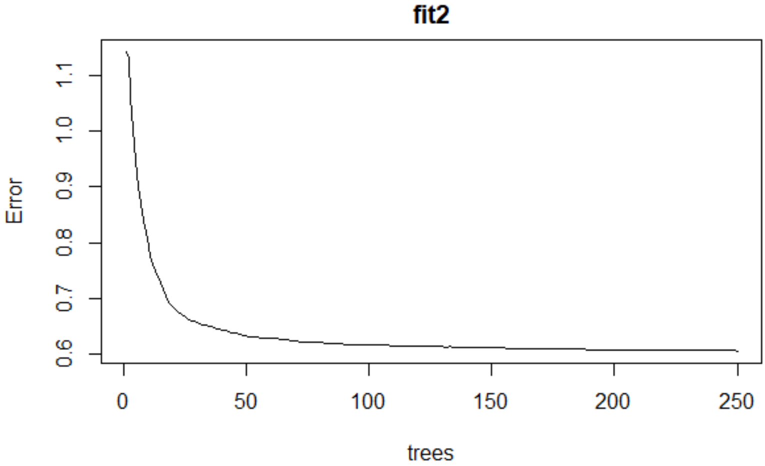
Fit1

Shows the relationship between tree number and error for our first model



Fit2

Shows the relationship between tree number and error for our second model



Wrap-Up

Ultimately this project established a strong relationship between weather (shown through temperature, humidity, and windspeed primarily) and bike rental demand. I was able to somewhat effectively predict demand, which allows policy makers, moving forward, to greater allocate resources in consideration of the demand structure.

The End