Capital Bikeshare Predictions

Joe Paolicelli

Who Am I?

I am a Senior at the George Washington University, majoring in Computer Science.

This past summer, I was a software development intern at Mission Data.

The Question

Can we use machine learning to predict the chance a bikeshare station will have a bike available, or a rack to check in a bike, at a future point in time?

Machine Learning Techniques

We used the [random forest](#) algorithm, after testing a number of options (including logistic regression, AdaBoost, and even some simple neural networks).

The features the algorithm learned from were:

- Time of Day
- Day of Week
- Month
- Temperature
- Amount of Precipitation

API

```
{

    "datetime": string (date and time),

    "location": string[,

    "station_count": integer]

}
```

API

- Station List
  - Name
  - Distance
  - Prediction
    - Chance Empty
    - Chance Full
- Forecast
  - Condition
  - Temperature (Celsius and Fahrenheit)
- Address, Date, Time

Data (and Sources)

Historical Bike Data - Capital Bikeshare Tracker

Current Bike Rack Status - Official Capital Bikeshare Station Status Feed

Historical Weather Data - NOAA

Weather Forecasts - Weather Underground API

Google Maps Geolocation API

Tech Stack

Machine Learning in Python 3, with Pandas and scikit-learn

API in Python 3 using Falcon

Web interface in Angular 2 with TypeScript, running on a Sinatra in Ruby
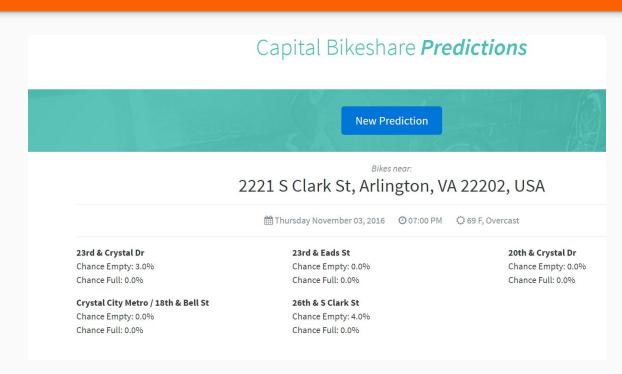
Results

Our predictions worked. On most bike stations, we were able to reach accuracy above 95%.

Cyclists don't like getting wet. Adding weather data improved results dramatically.

Interfaces

A public web application.

A conversational Slackbot using wit.ai.

Links

Mission Data Lab Notes: [Machine Learning Part 1](#), [Part 2](#), [Slackbot](#)

[Source (on Github)](#)

[Mission Data](#)

Thank You!