

# Bike share traffic predictions

using **machine learning**

Arnab Kumar Datta

UiO : **Department of Informatics**  
University of Oslo





# Agenda

- Introduction to bike-sharing
- Motivation and vision
- A short introduction to machine learning
- Overview of software
- Results
- Conclusion



# Bike-sharing

Why bike-sharing

The problem

[< Back](#)

Reviews

[Write a Review](#)

## "Great fun but..!"

●●●●○ Aug 28, 2013



Fatima M (1 Review)

London, United Kingdom

We hired the bikes for 4 people and had an amazing time! The only catch is that sometimes (specially over weekends) when you want to place the bikes back into their docking stations, you might not be able to get the space. We tried three different stations but ~~couldn't find a single empty space and were thoroughly exhausted after the whole day.~~ Because we were stationed in london for the night at a friends' place, we had to park the bikes in their apartment overnight and only in the morning did we find empty spaces in the nearest docking stations. So great fun and highly recommended but you might want to get back to the docking stations in, shall I say a "lucky" time :)

Above: A customer reviews London's bike-share system on the tripadvisor website



duffa101  
Oxford

**Senior Contributor**



33 reviews



17 attraction reviews



Reviews in 15 cities



24 helpful votes

***"A good way to get around - but don't expect to find a bike every time."***



Reviewed July 30, 2013

First off: hats off to any city that's doing its bit to promote this kind of transport system. Washington's heart is certainly in the right place. \$7 for a day of travel around the sights of the National Mall and further afield to Arlington and Georgetown is certainly good value. To clarify: you have 30 minute window to get your bike docked into another station, not the 90 as one review has it, though you do get a 'safety' 15 minutes if you find all of the stations full when you reach your destination. Given that the metro is \$2.70 a ride (when including the paper card) you can save yourself some money and sore feet with the bike.

The bikes are simply and easy to ride, and you can get a reasonable shunt going if you're so inclined, though I'm not sure why you'd want to unless you're a DC commuter heading into work, and then why are you reading this?

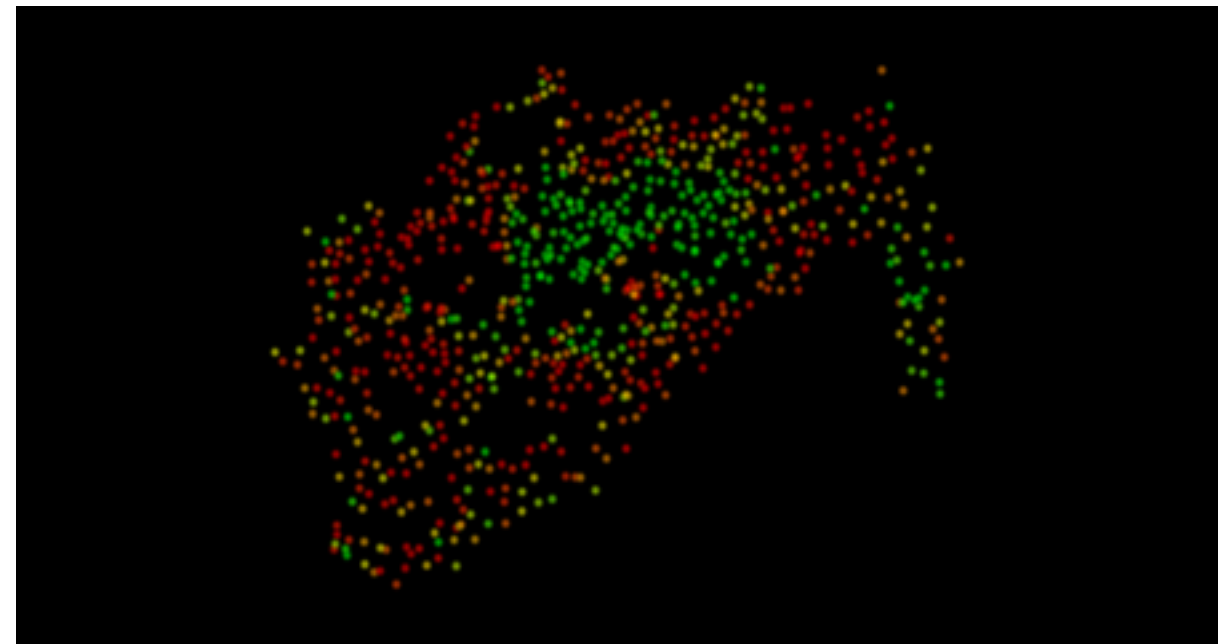
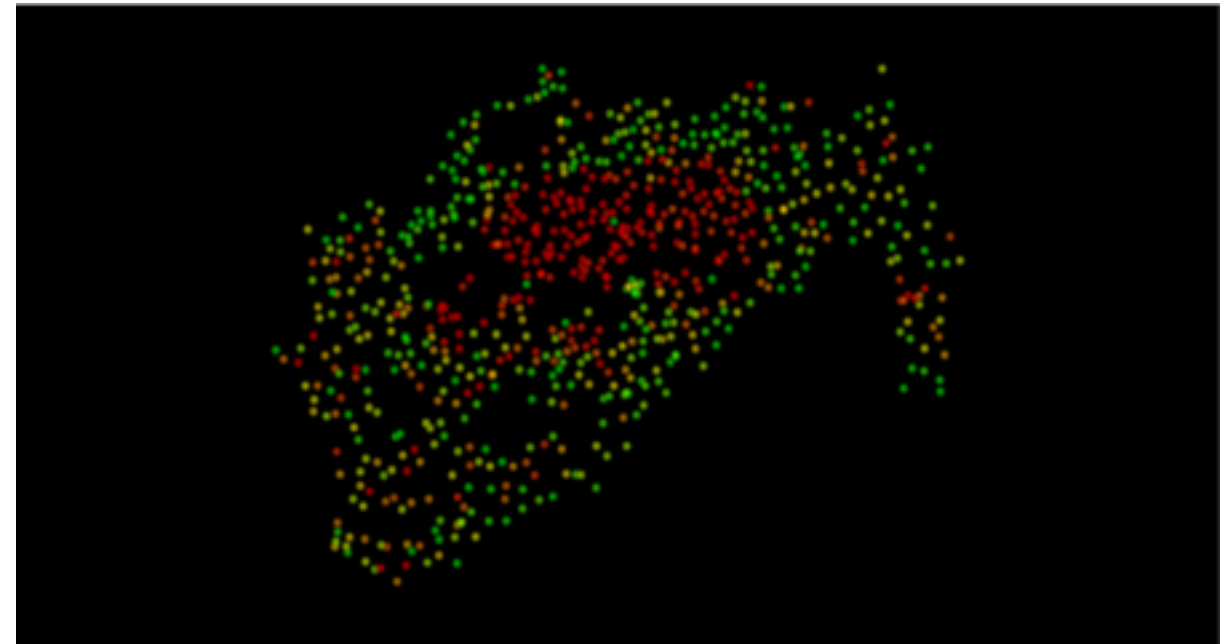
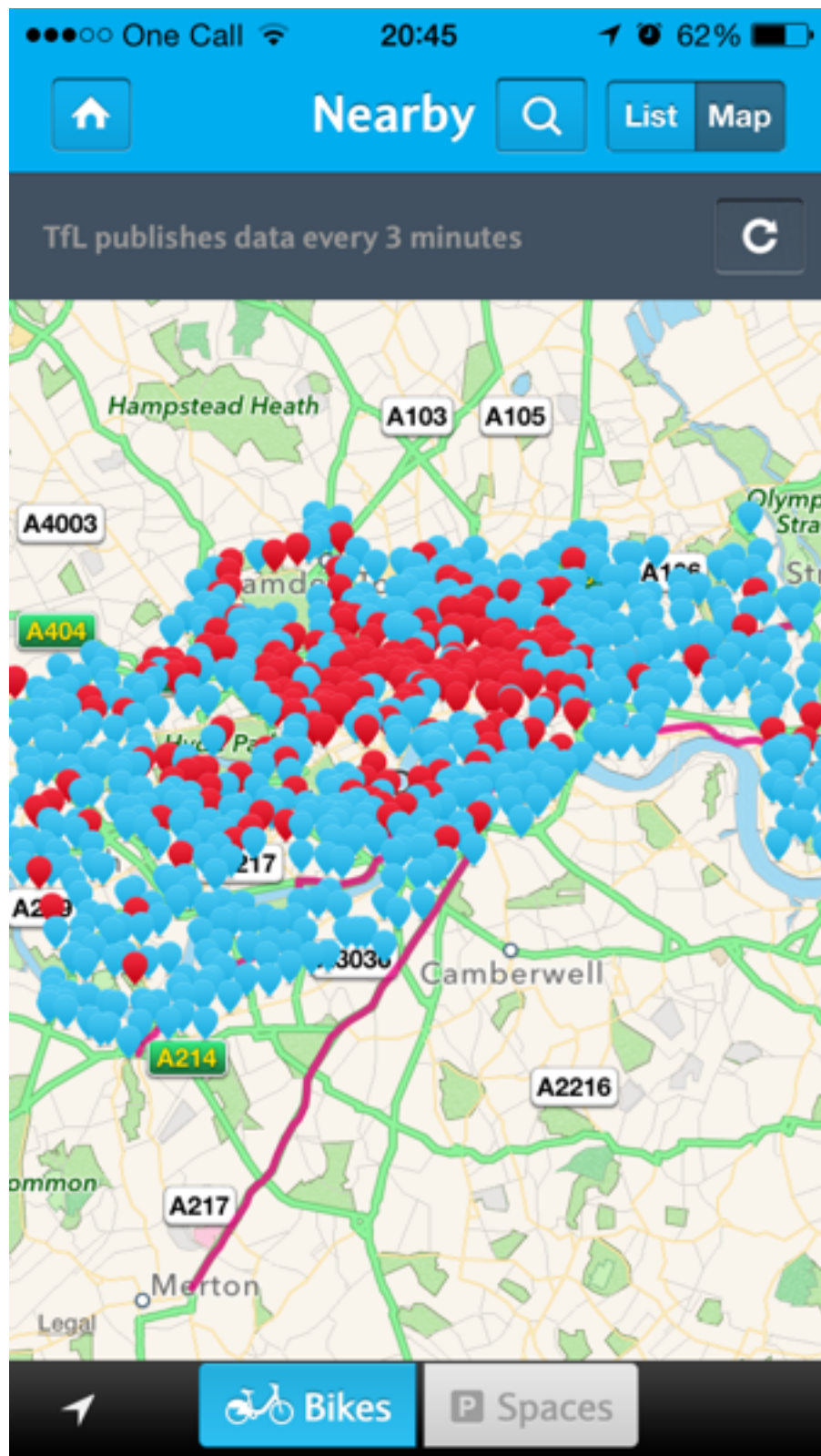
Here's the health warning: there are lots of tourists and bikes can become a premium in certain places (the Lincoln Memorial for example) so you can't expect to rock up and find one. The app is good, but doesn't update strictly in real time: by the time you reach the station the bikes might be gone. Some people are quite canny at dropping a bike then swiping for another 30 minutes, so pairs of bikes are sometimes hard to come by.

Personally, I wouldn't plan a day around using a bike, especially if you have are under time pressure to get back somewhere, but if you can afford to be a bit more relaxed this might be a fun, and useful, bonus attraction.

**Above: A customer reviews Washington's bike-share system on the tripadvisor website**



Users currently have real-time systems





# The vision

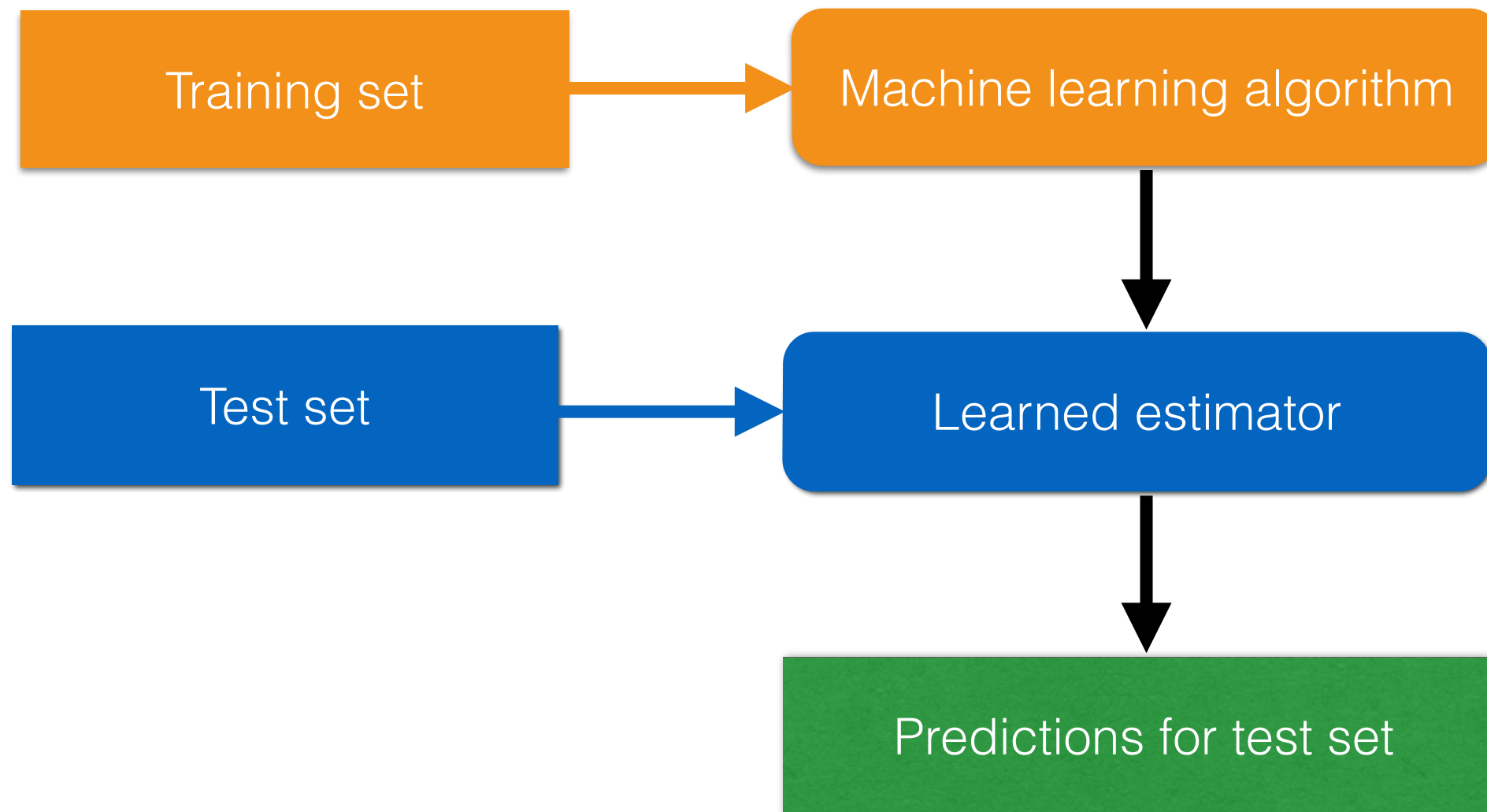
**"I will be downtown at 8 am on Monday. Will the bike station be full?"**



# Related work

- Data science for social good  
(predicting bike-share usage in Chicago's Divvy bike system)
- Jake VanderPlas (modelling the effects of weather on bike usage in Seattle)

# Machine learning





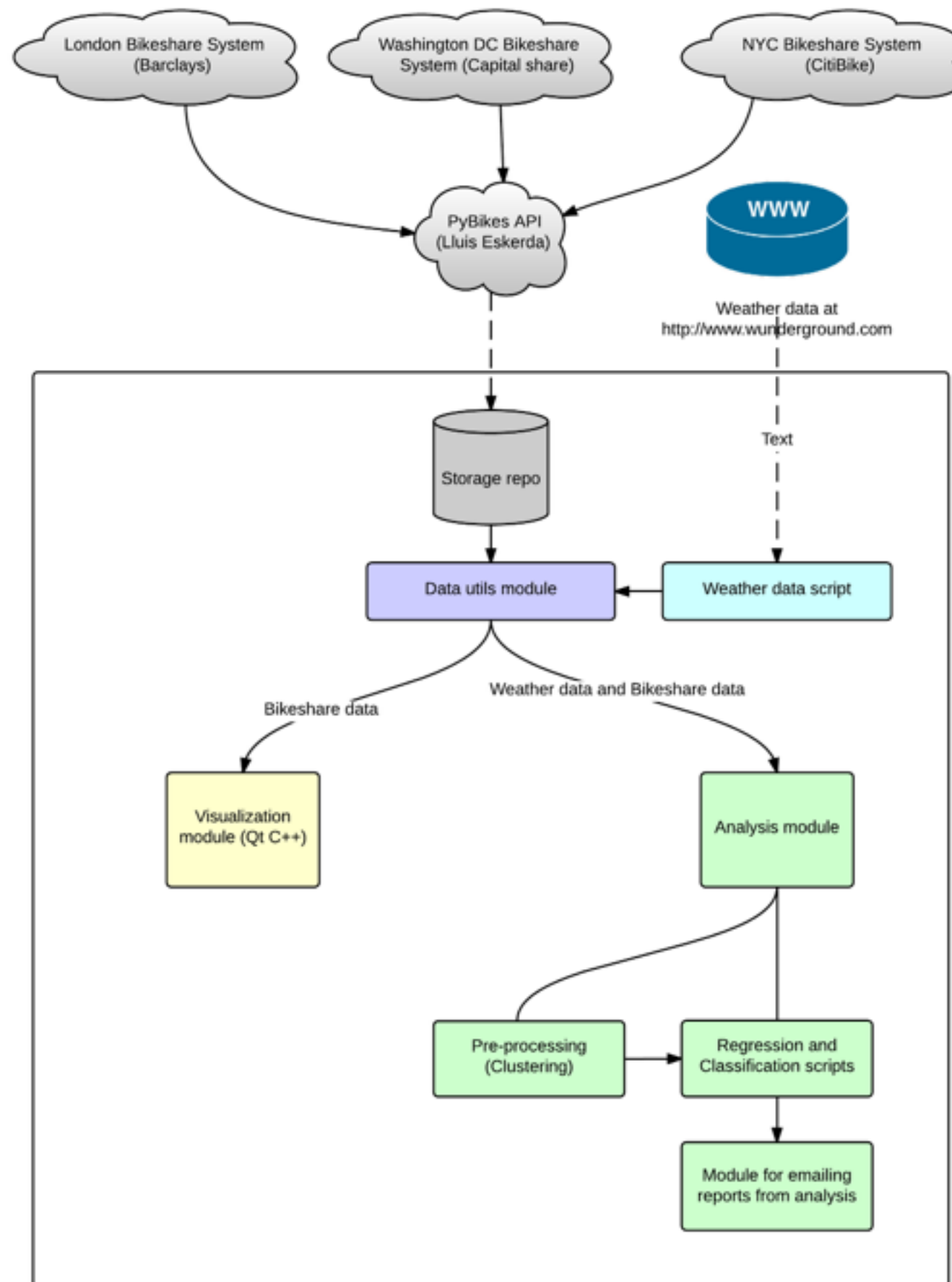
# Training set

Sunny	Downtown	Tuesday 8:00 AM	11 bikes
Sunny	Downtown	Tuesday 11:00 AM	0 bikes
Rainy	Downtown	Tuesday 8:00 AM	2 bikes
Sunny	Downtown	Tuesday 11:00 AM	2 bikes
Sunny	Downtown	Tuesday 1:00 PM	1 bike

# Test set

Sunny	Downtown	Tuesday 8:00 AM	11 bikes
Sunny	Downtown	Tuesday 11:00 AM	1 bike
Sunny	Downtown	Tuesday 8:00 AM	10 bikes
Sunny	Downtown	Tuesday 1:00 PM	2 bikes
Sunny	Downtown	Tuesday 2:00 PM	1 bike

# Software overview



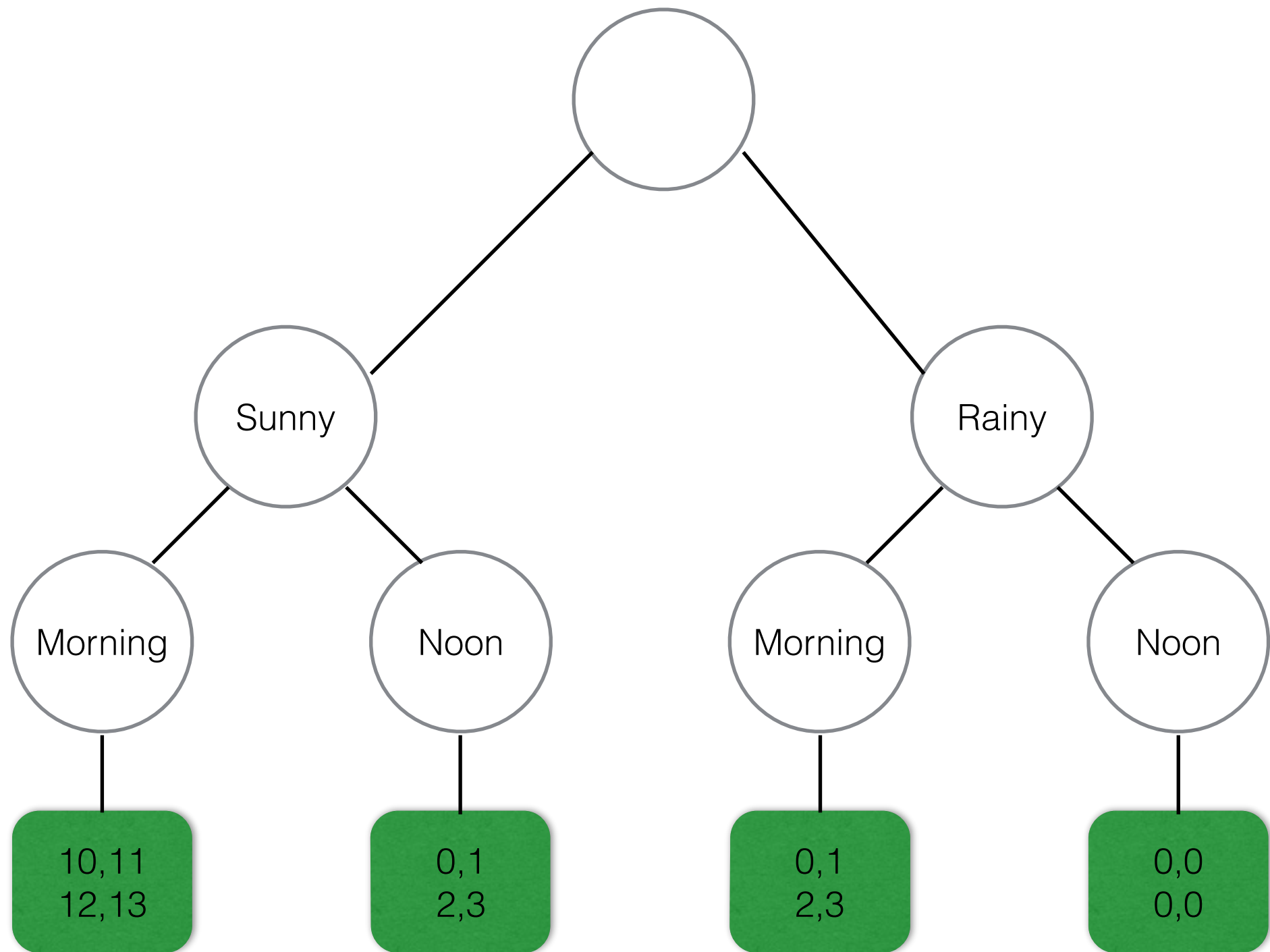


# Libraries used

- Scikit-learn (machine learning algorithms)
- Pybikes (data collection) to collect data from the Washington bike-share system

# Machine learning algorithms

# Decision Trees





# Random Forests

# Random Forests

- Lots of decision trees
- Output given by the average of the output of all trees in the forest
- Cannot overfit by adding more trees (note: RF can overfit on noisy datasets when there are too few trees!)

Ada Boost

# AdaBoost

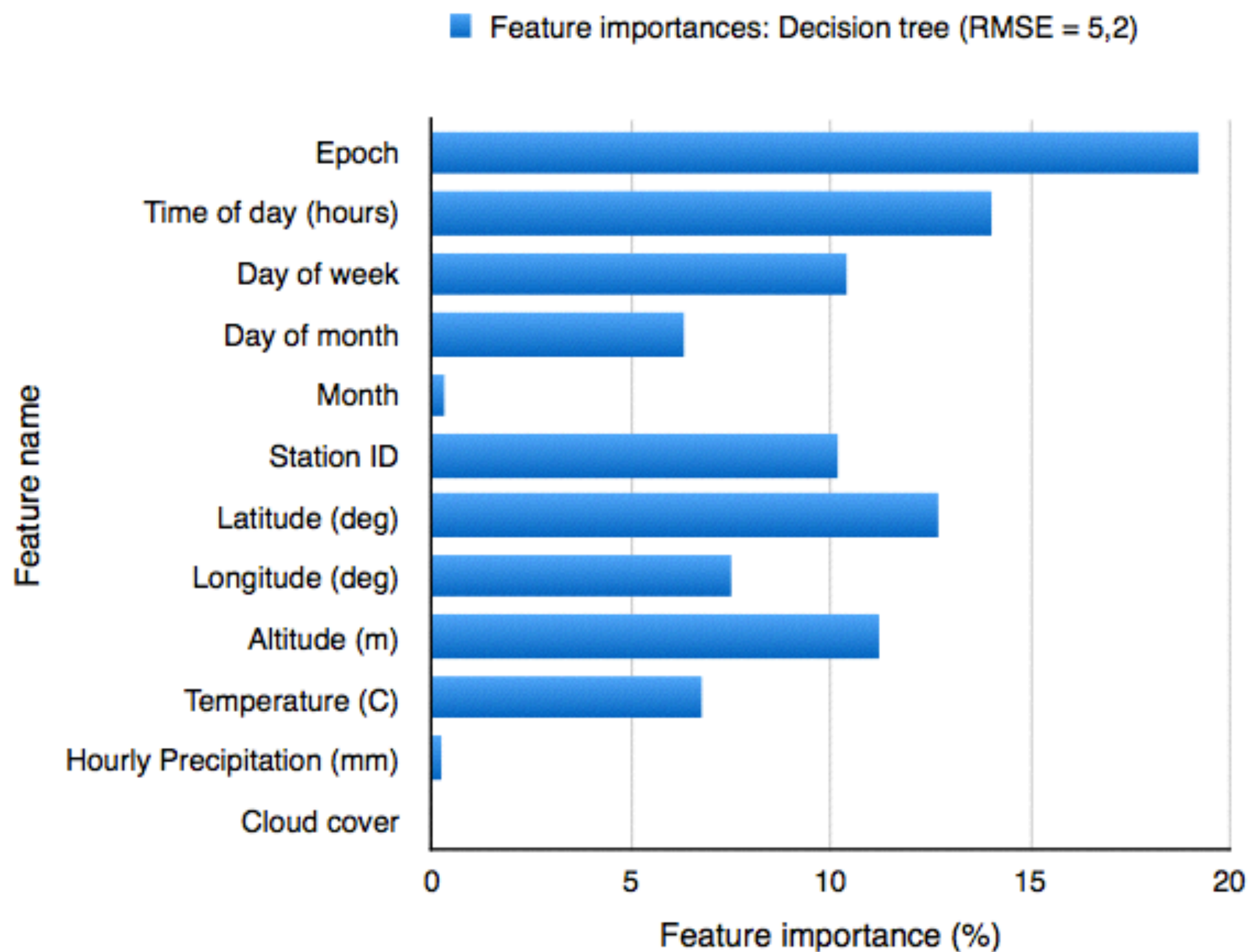
- Analogy: student preparing for an exam in physics
  - Topics covered: classical physics, thermodynamics, electromagnetism, quantum physics
  - They start by doing a practice exam
  - They notice they didn't do well on electromagnetism. Ignore all other topics until they grasp electromagnetism.
  - Do another practice exam
  - Repeat... until it's time for the exam

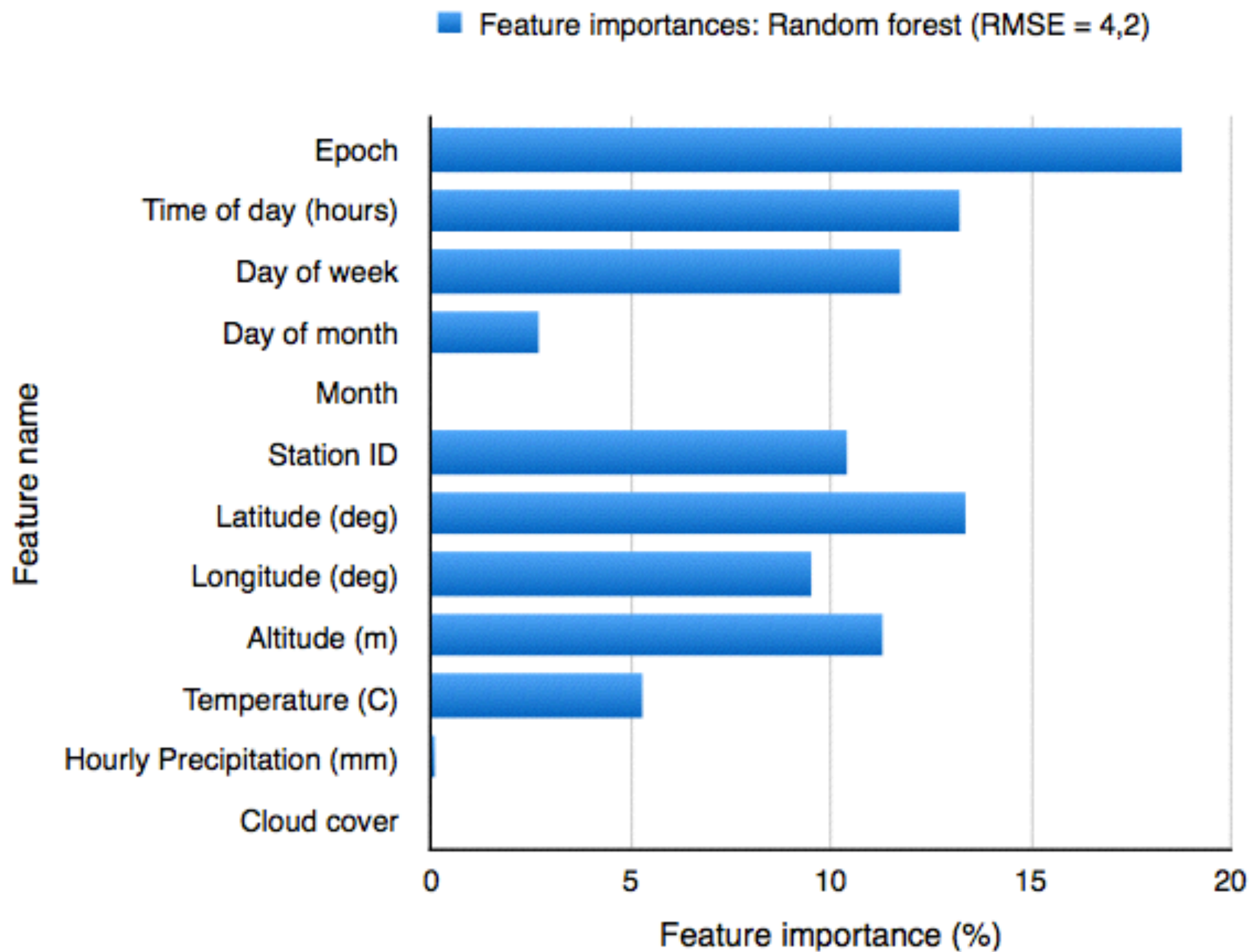


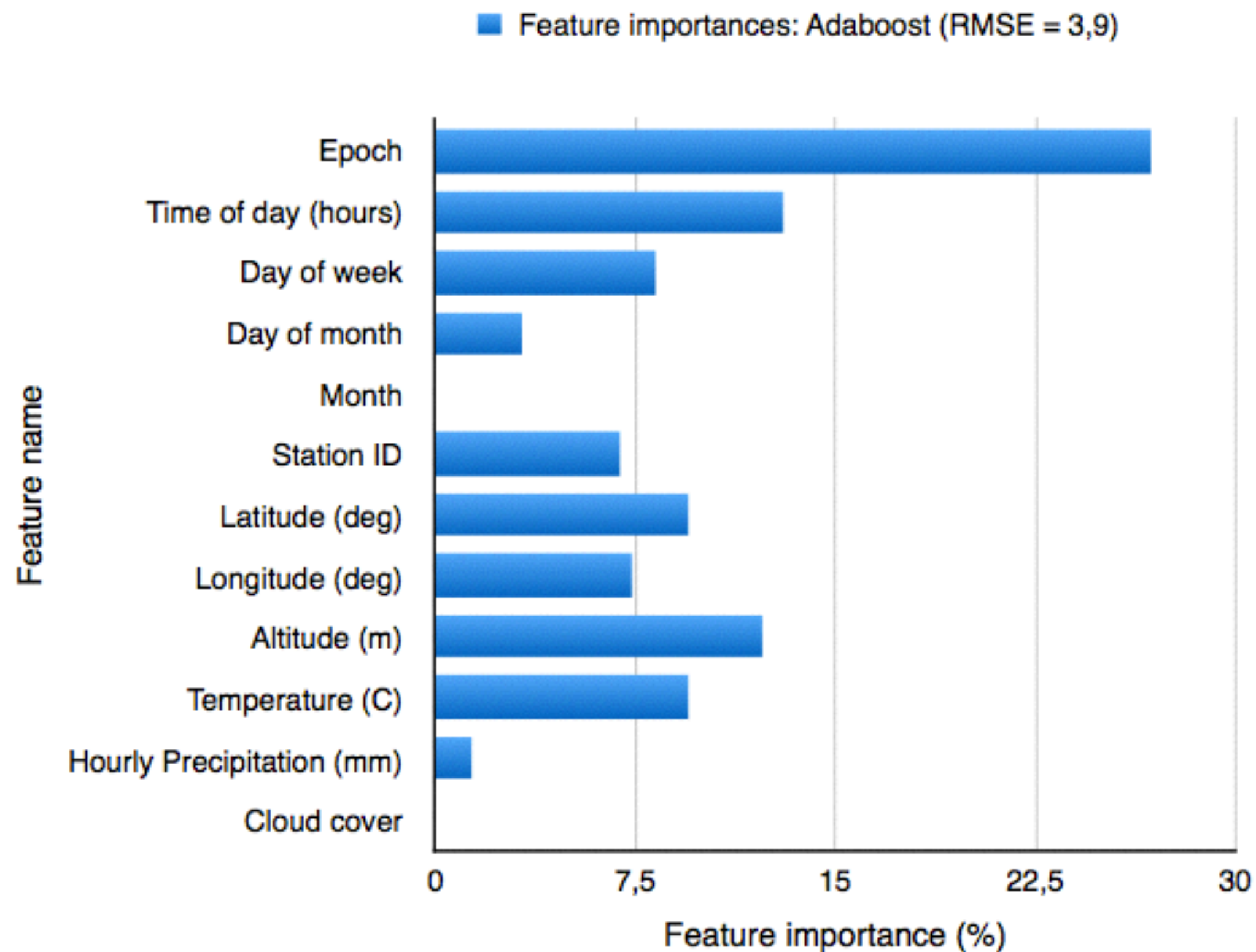
# Thesis contribution

# Data collection using Pybikes

# Feature selection







# Why is the “epoch” so important?

A missing time-related feature that has not been accounted for.



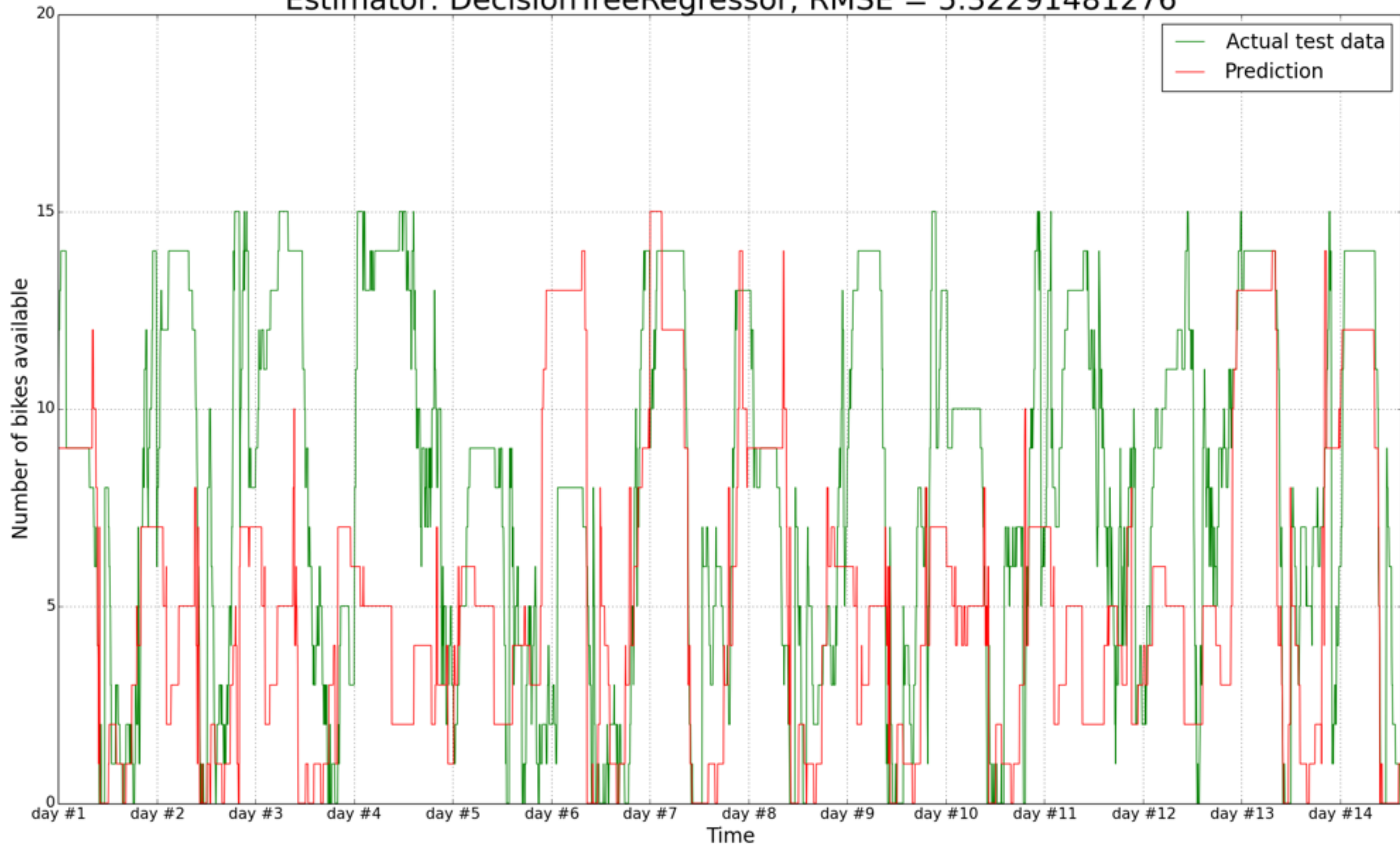
# Genetic algorithms

- Hyperparameters - algorithm configuration
- Can use GA to pick the “optimal” feature set that provides the best prediction performance
- GAs did not improve the accuracy over manually picked hyperparameters

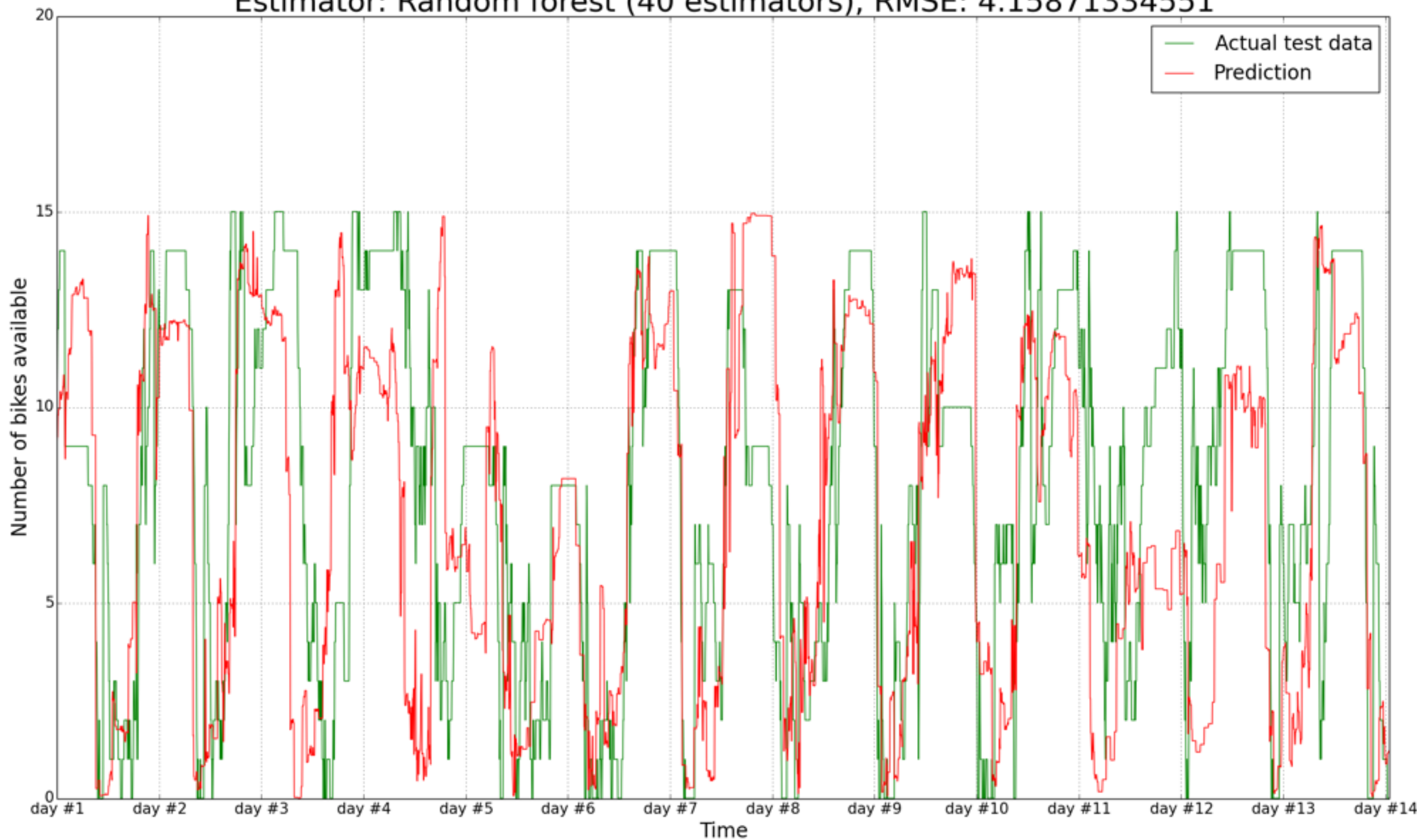
# Results

A customizable  
machine-learning  
package for predicting  
bike-share usage

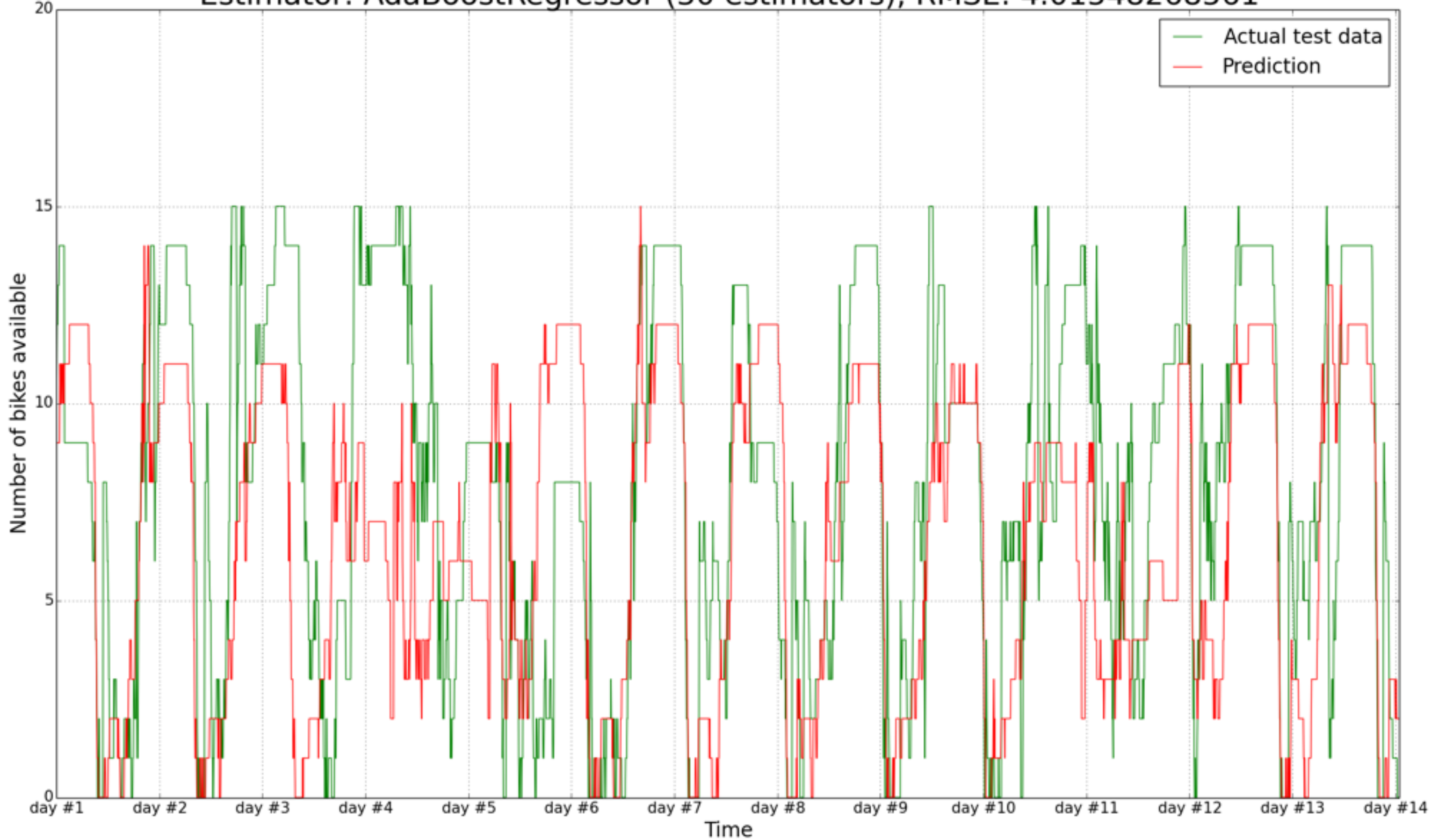
Estimator: DecisionTreeRegressor, RMSE = 5.32291481276



Estimator: Random forest (40 estimators), RMSE: 4.15871334551



Estimator: AdaBoostRegressor (30 estimators), RMSE: 4.01548268561

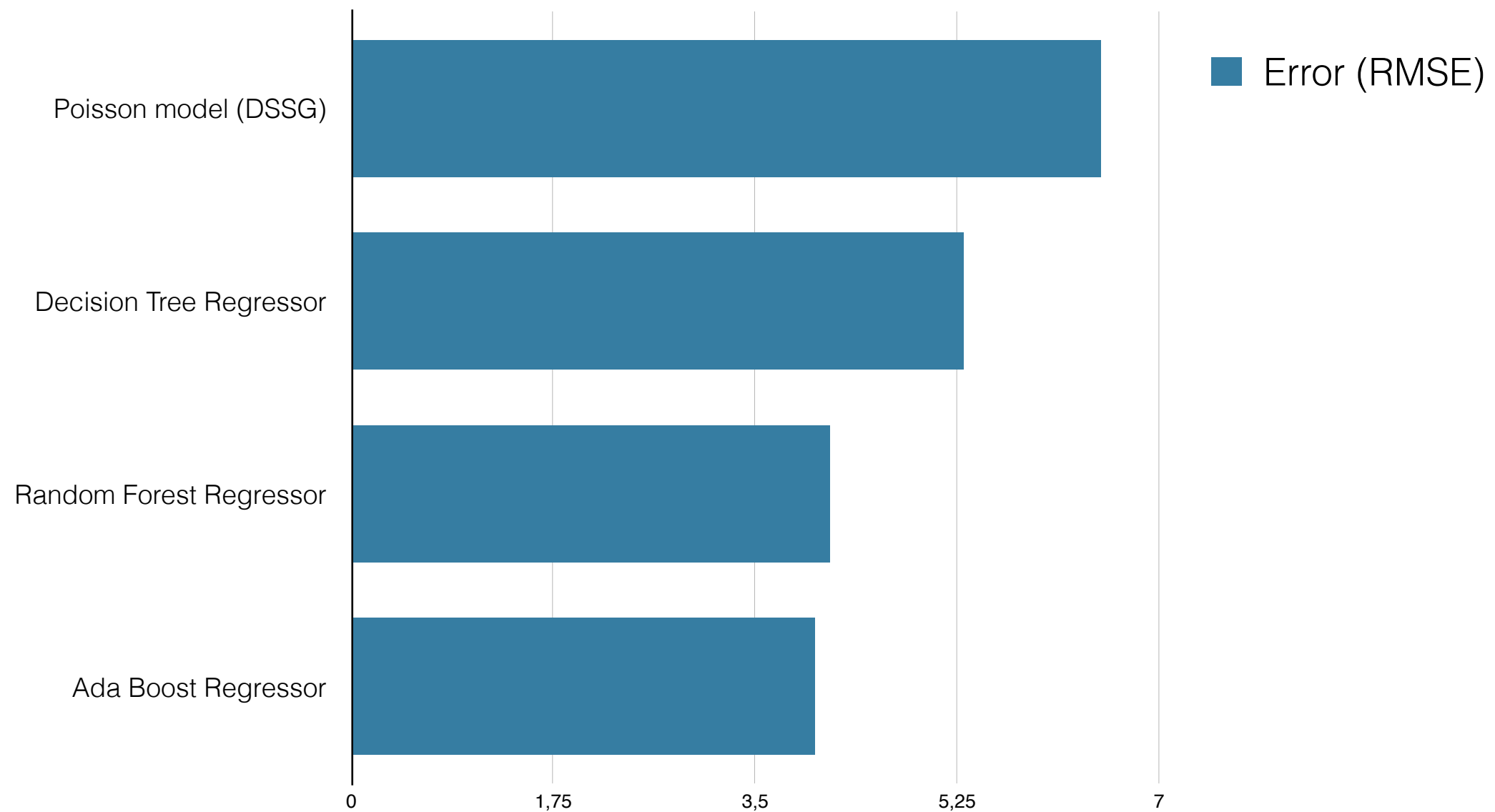


Improvements on  
existing solutions?

Error metric: RMSE



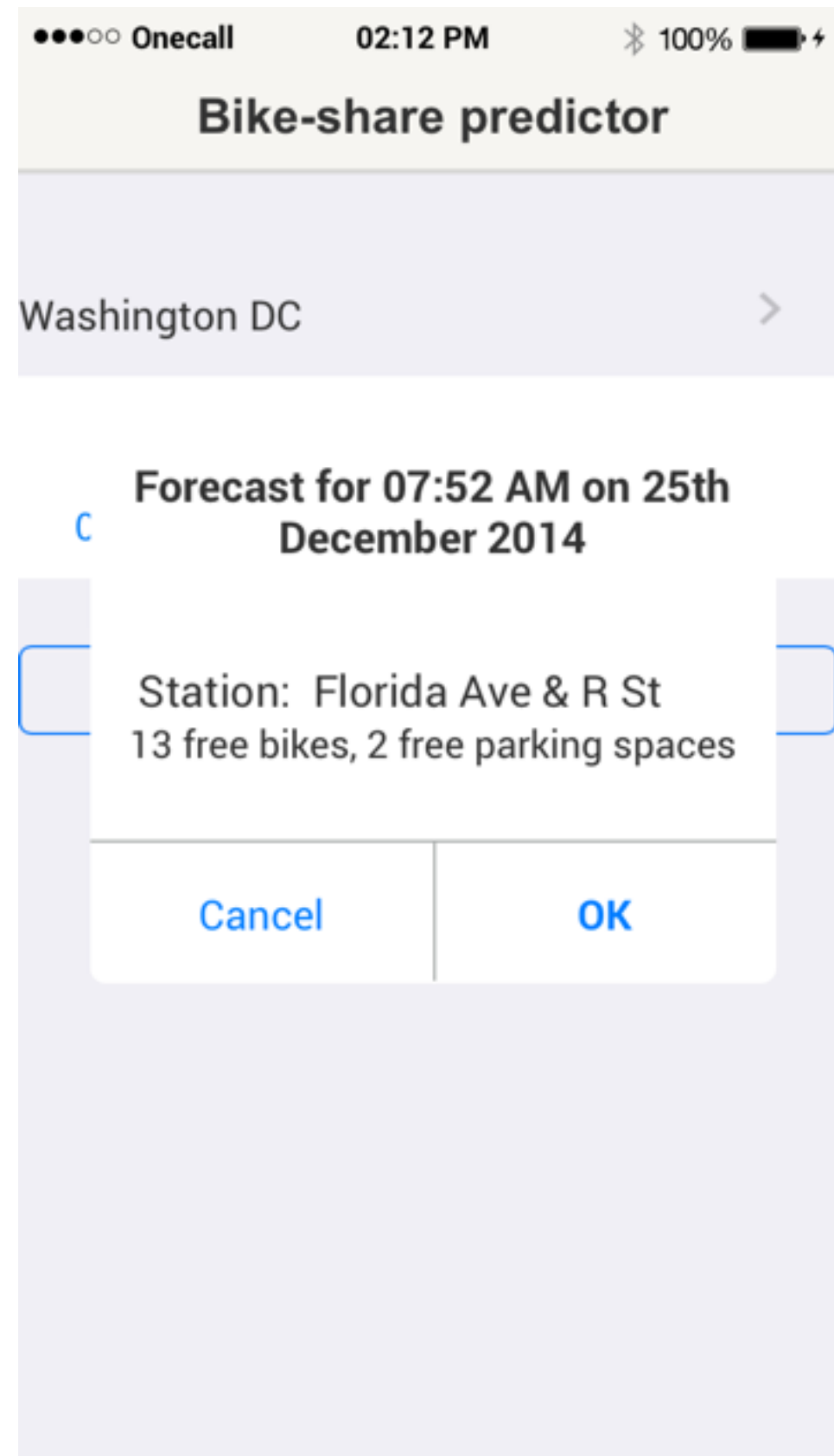
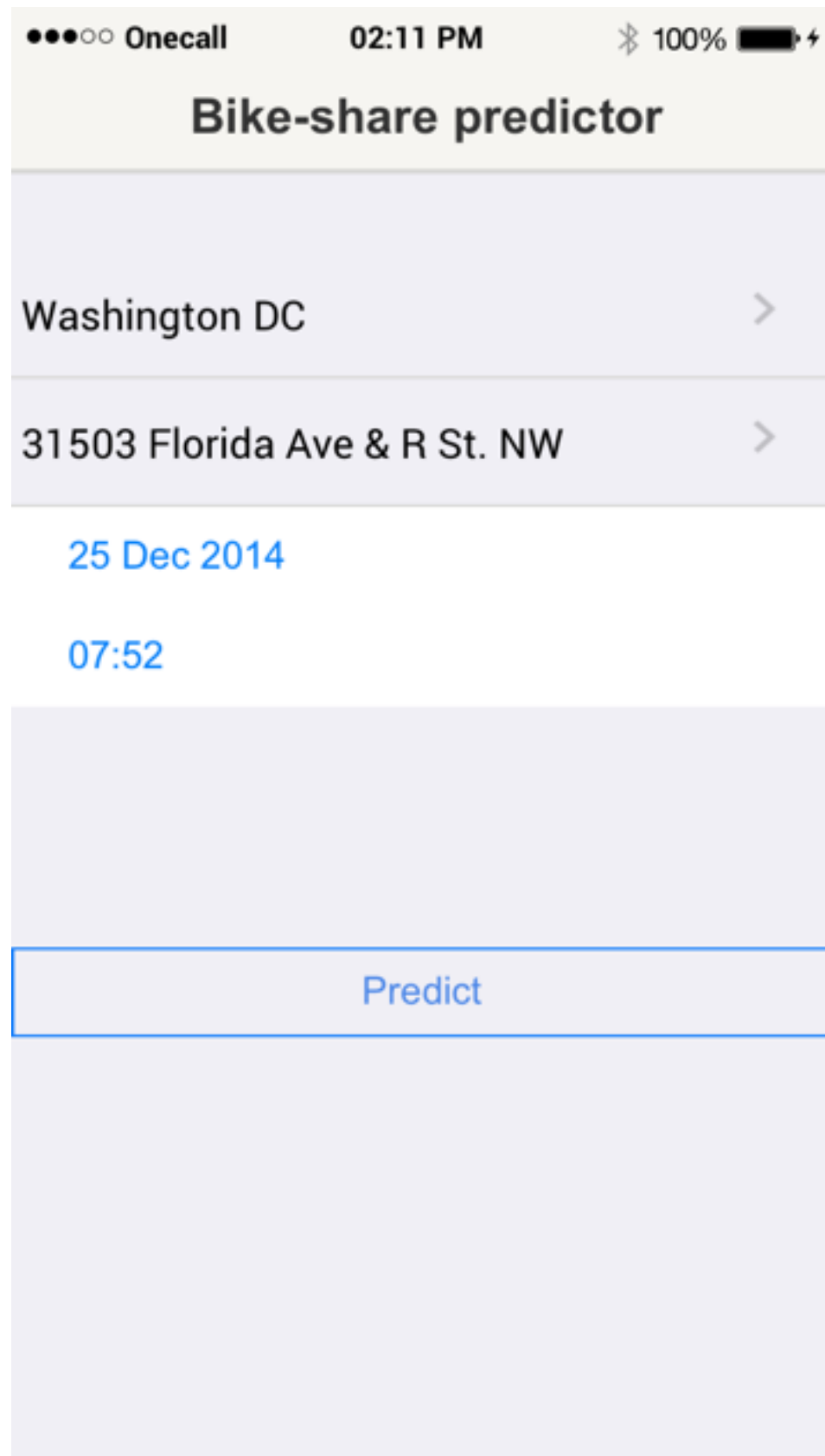
# Improvement



Further work

# The vision





Settings

Predictions

Alerts (4)

Prediction date(s)

10.12.2014 - 11.12.2...

Weather Forecast API

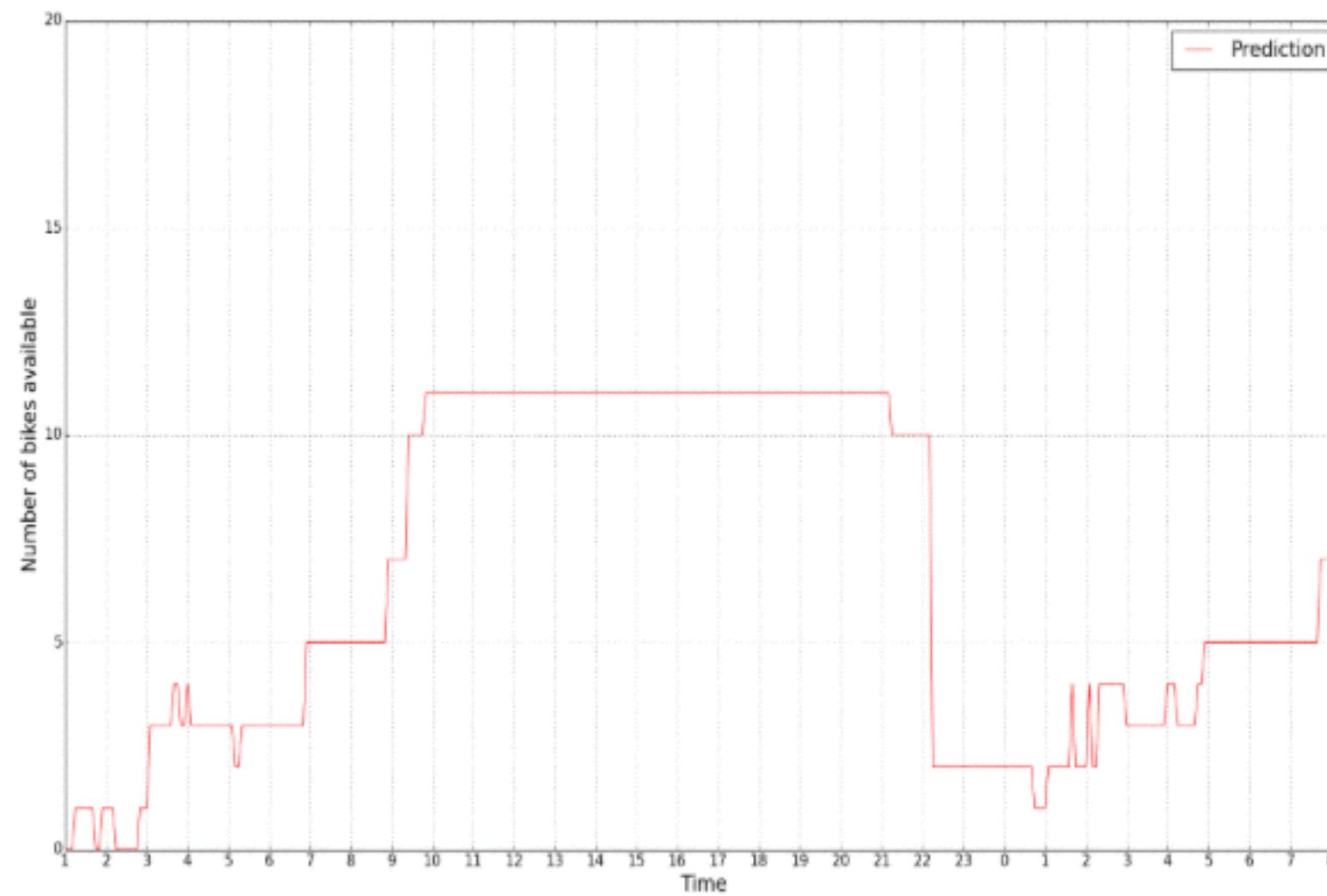
www.wunderground.com

Bike-share system

Washington DC

Station to predict

59: Florida Ave & R St.



---

# Bike-share traffic prediction

---

[Settings](#)

[Predictions](#)

[Alerts \(4\)](#)

Stations predicted to be empty:

73. L'Enfant Plaza / 7th & C St SW (in 30 mins)  
78. Tenleytown / Wisconsin Ave & Albemarle St Nw (in 15 mins)

Stations predicted to be full

95. Bladensburg Rd & Bening Rd NE (in 140 mins)  
55. Potomac & Pennsylvania Ave SE (in 30 mins)

---

# Bike-share traffic prediction

Settings

Predictions

Prediction date(s)

10.12.2014 - 11.12.2...



Weather Forecast API

www.wunderground.com/

Bike-share system

Washington DC



Station to predict

59: Florida Ave & R St.

