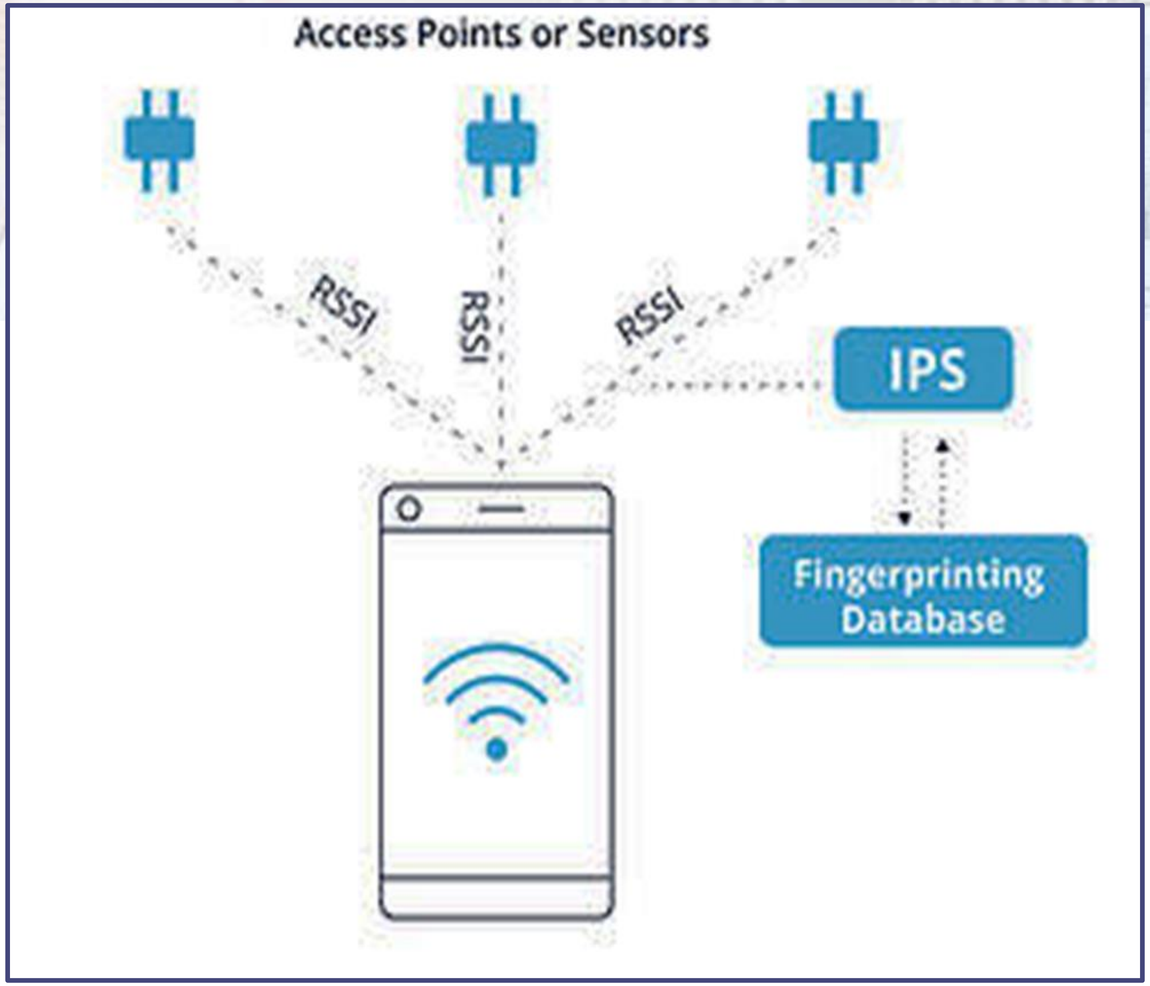
LOCATIONING WITH WI-FI FINGERPRINTING

Kate Koebbe

WHAT IS WI-FI FINGERPRINTING?

In the absence of strong GPS signal indoors, new techniques are being employed in locationing to overcome that challenge. Wi-Fi fingerprinting is one such technique, using signal strength measurements from wireless access points (WAP) to Wi-Fi enabled devices to determine position.



REAL WORLD APPLICATIONS

Emergency Response

Responders locating individuals in crisis situations more efficiently.

Targeted Advertising

Customers receive notifications of promotions based upon location in a store.

Smart Device Utilization

Smart home device automations triggered when specific proximity thresholds are met.

“BUT DOES IT WORK?”

STUDY GOALS

Our team set out to investigate the feasibility of using wi-fi fingerprinting to determine a location in indoor spaces by evaluating multiple machine learning models to see which produces the best result.

DATA MANAGEMENT

Gather & Evaluate

Analyze & Visualize

Model & Predict

PREPROCESSING NOTES & METHODOLOGY

Making Our Study More Efficient

- We performed an initial removal of all columns that did not provide location information (USER ID, PHONEID, TIMESTAMP).
- We also removed LATITUDE and LONGITUDE information because, while precise, does not give us corresponding information inside of the building.
- We created a unique attribute, “LocID,” which is a number that combines both FLOOR and SPACEID information. For example, if a LocID is 1105, the location is Floor 1, Space 105.

PREPROCESSING NOTES & METHODOLOGY

Making Our Study More Efficient

- We subset by BUILDINGID (in our study, Building 0) to reduce the number of observations processed by our algorithms, thereby making the processing faster. When we select the best fit algorithm for our subset, we can apply the other subsets to make predictions.

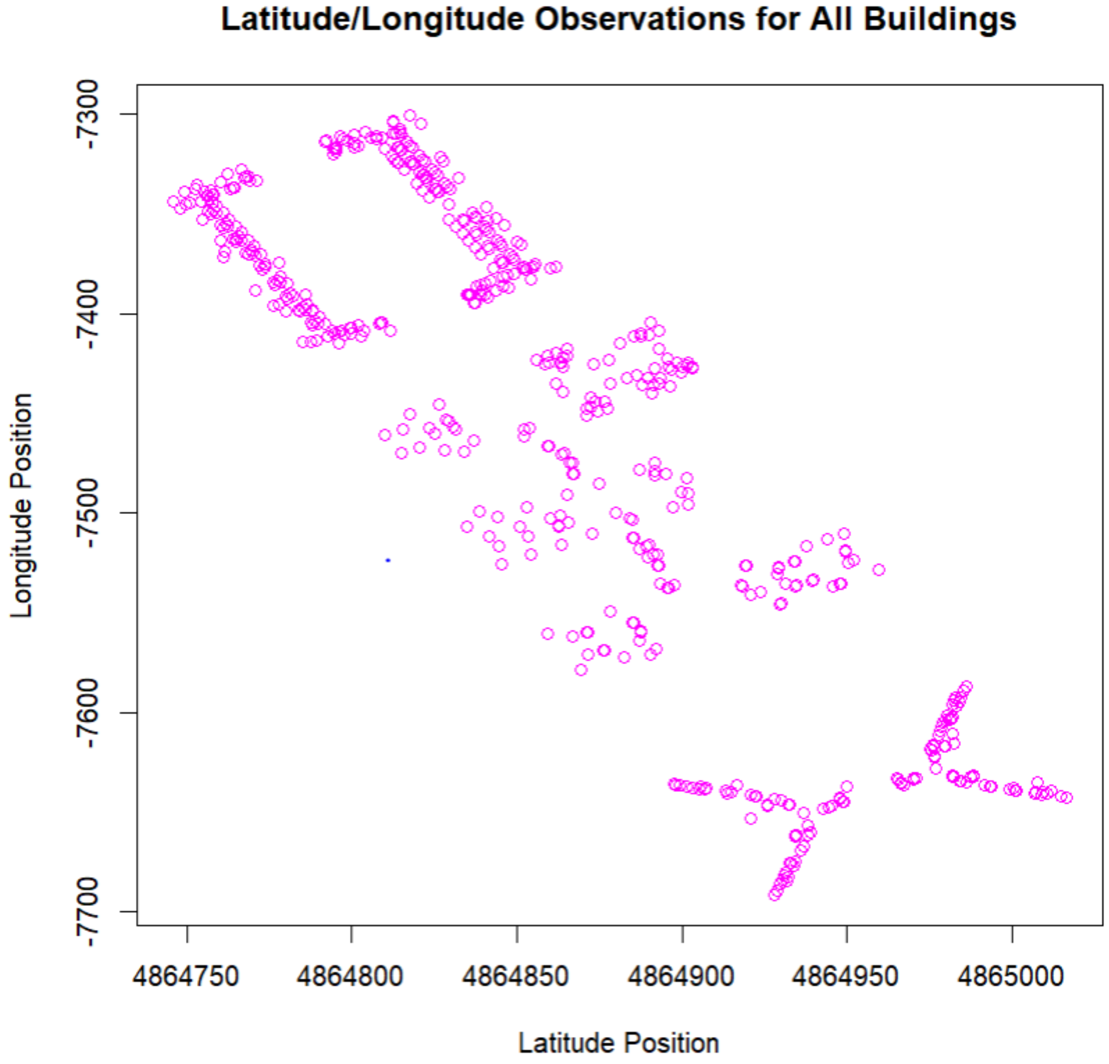
- We tested four classification models to see which performs the best with our dataset: Naive-Bayes, K-Nearest Neighbors, C5.0, and Random Forest.

- Finally, our principal measures of performance will be *accuracy* (number of instances that were classified correctly) and *kappa* (a comparison of observed accuracy and expected accuracy).

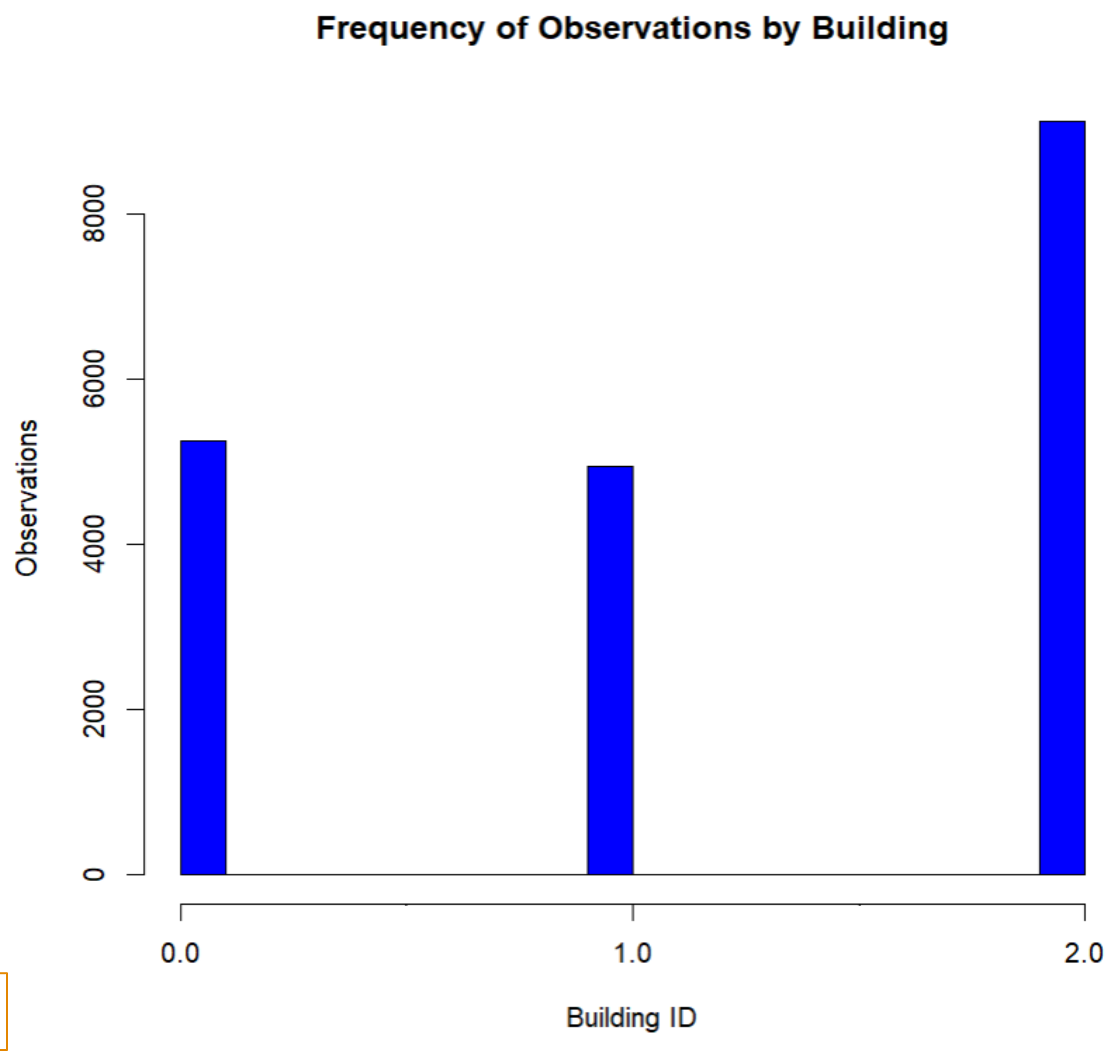
OUR DATA

INTRODUCTION TO THE DATA SET

- 19,937 observations from 529 variables.
- Wi-fi access points (WAPs) make up 520 of the variables in the data set.
- The WAPs are located across three buildings.
- Spatial variables captured: Latitude, Longitude, Building ID, Floor ID, Space ID, Relative Position.
- User ID, Phone ID, and Timestamp were also recorded but excluded from analysis as they do not offer insight into determining location.
- At right, the scatterplot of latitudinal and longitudinal observations gives us an understanding of the layout of the buildings.



RECORDED OBSERVATIONS BY BUILDING

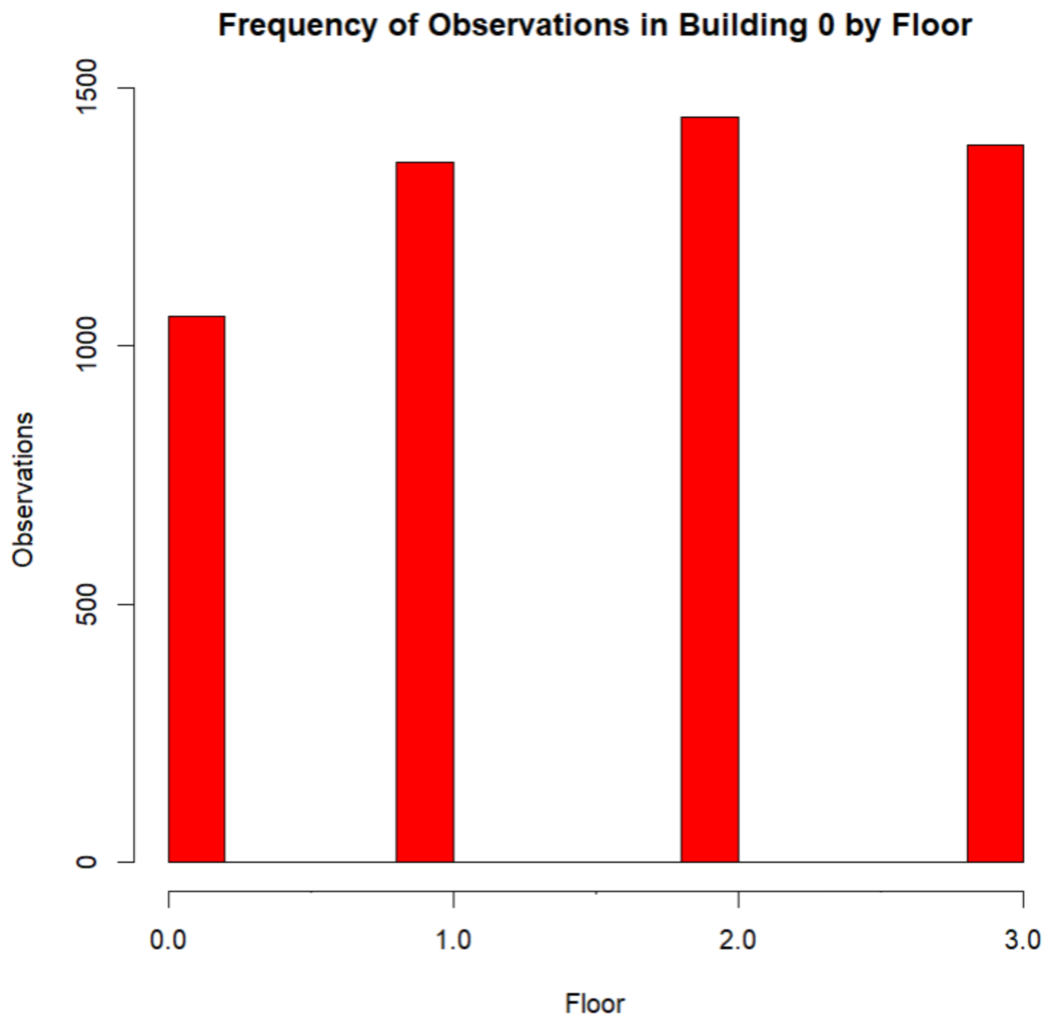


- Across all of the buildings, Building 2 has the most observations (n=9,116). It also has one additional floor.

- Buildings 0 and 1 have a similar frequency of observations.
    - Building 0 = 5,246
    - Building 1 = 4,938

RECORDED OBSERVATIONS BY FLOOR

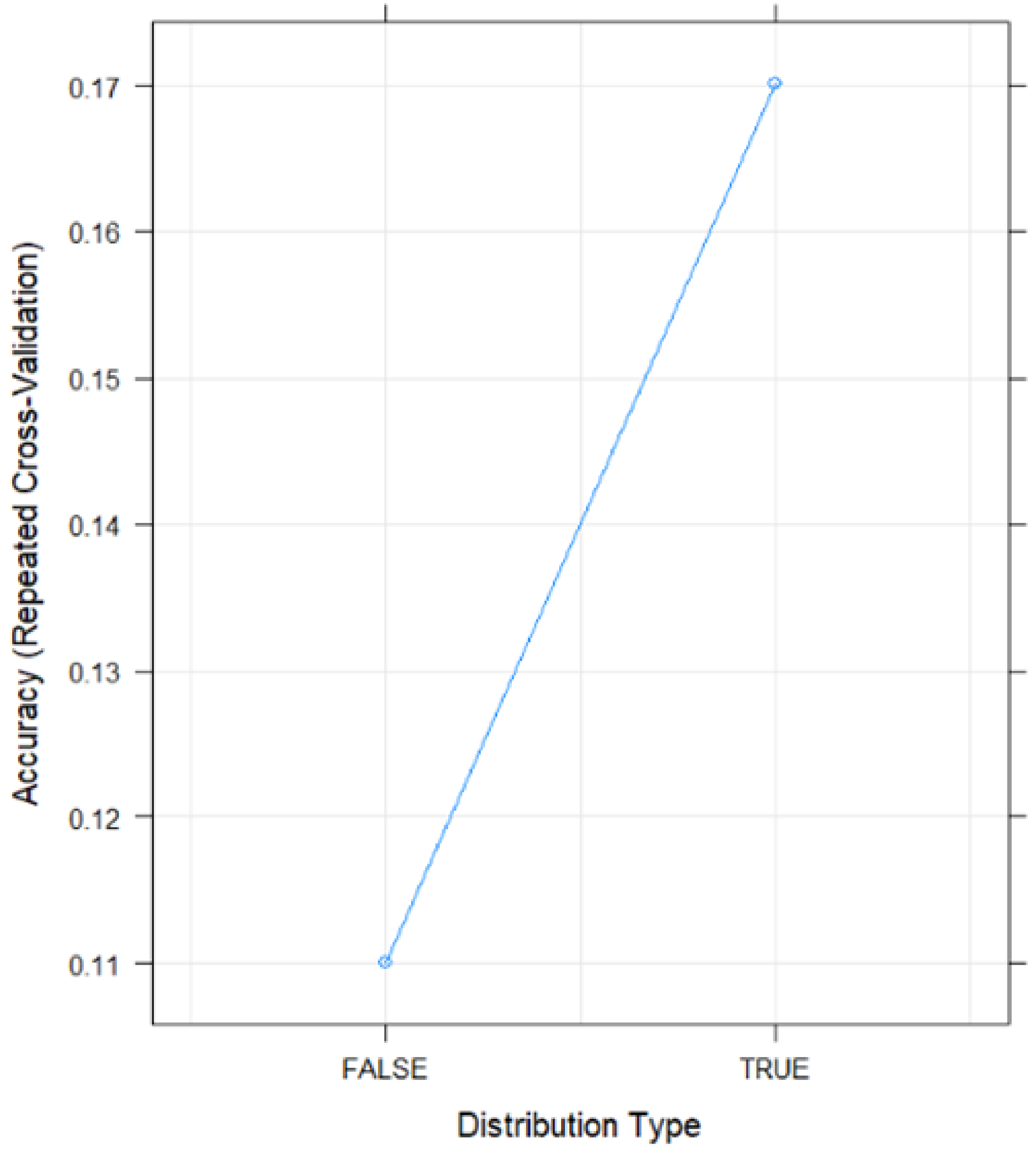
- At right, we can see the number of observations in a target building (Building 0).
- Floor 2 has the highest number of observations (n=1,443).
- Floors 1 and 3 have a similar frequency of observations (n=1,356 and n=1,390 respectively).
- Floor 0 has the fewest observations (n=1,057).



MODELING

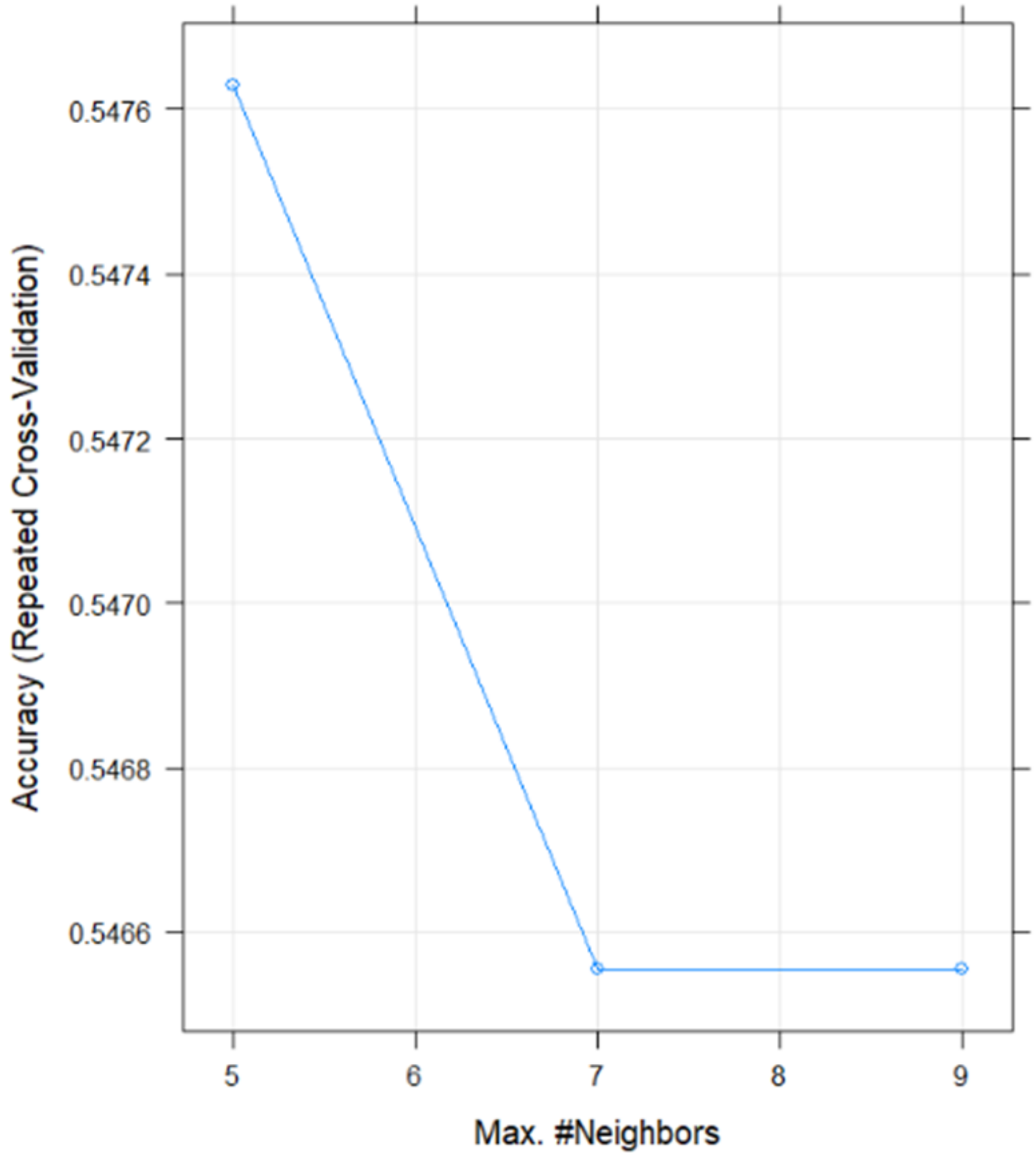
NAIVE BAYES

- Accuracy: 0.1701336
  Kappa: 0.1652163
- Naive Bayes was our lowest performing model for this data.



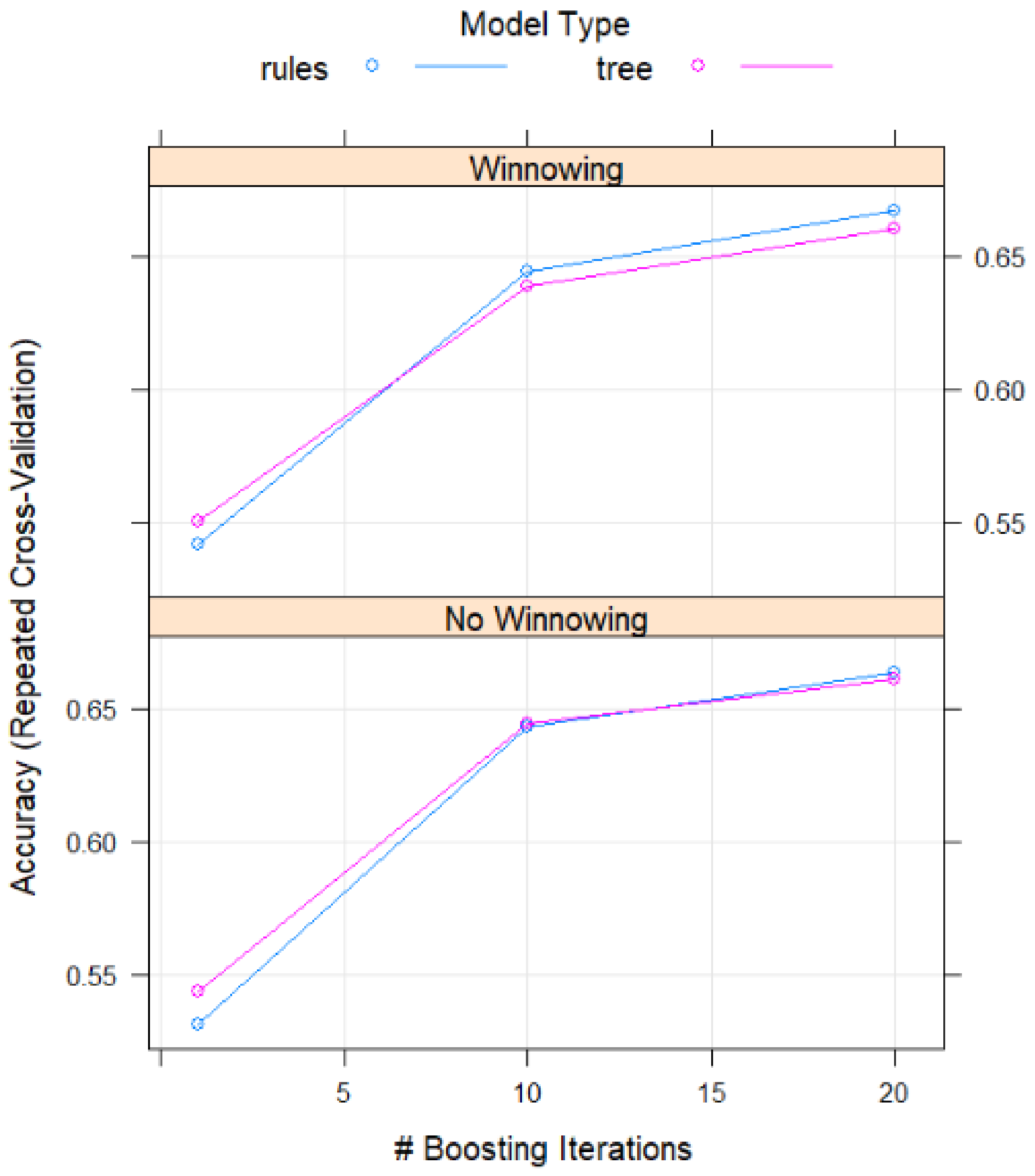
K-NEAREST NEIGHBORS

- Accuracy: 0.5476298 Kappa: 0.5457635
- A significant improvement over Naive Bayes but still our second lowest performing model.



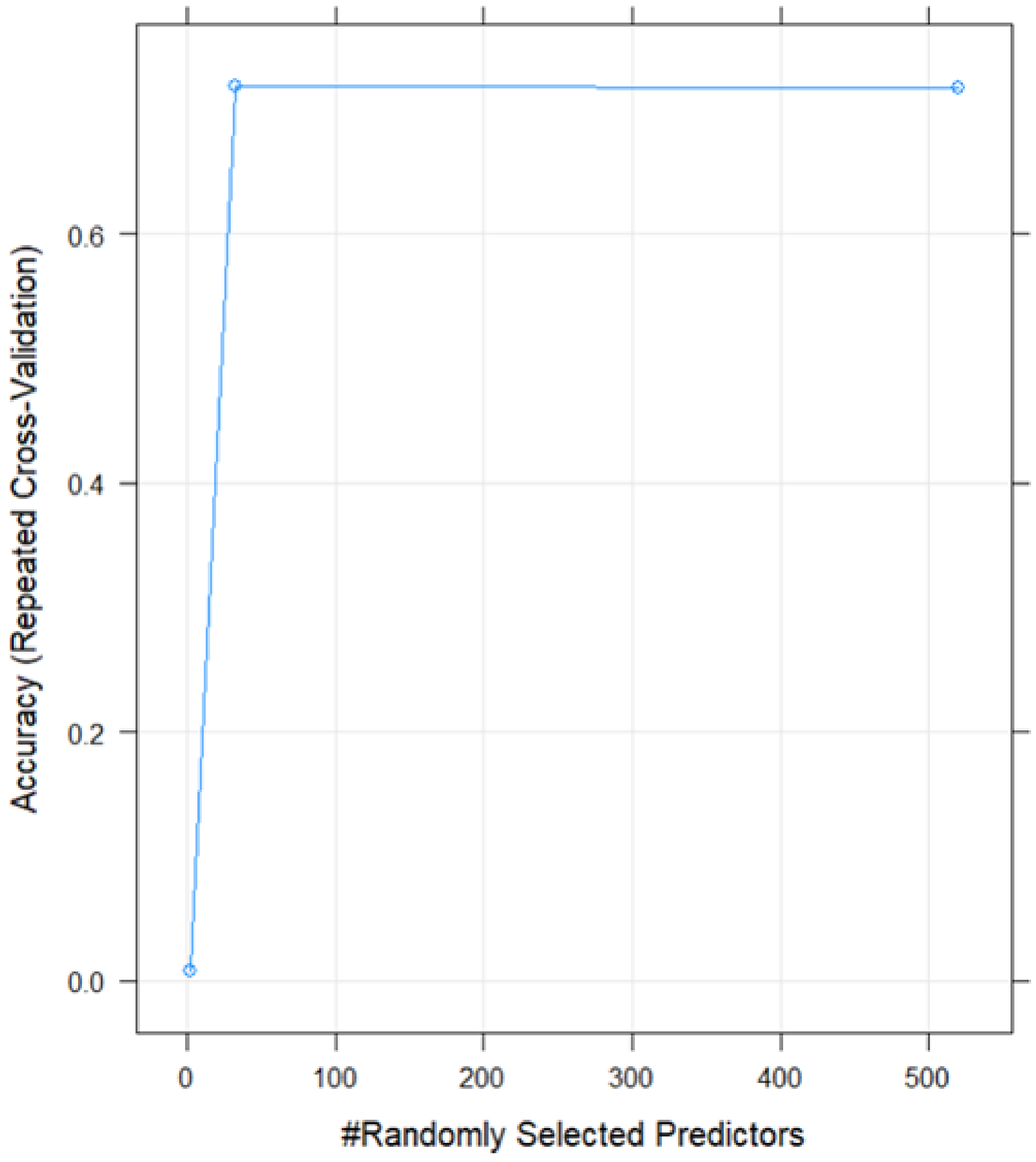
C5.0

- Accuracy: 0.6670973 Kappa: 0.6657256
- Better performance than Naive Bayes and K-Nearest Neighbors but not as good a fit to the data as we would like.



RANDOM FOREST

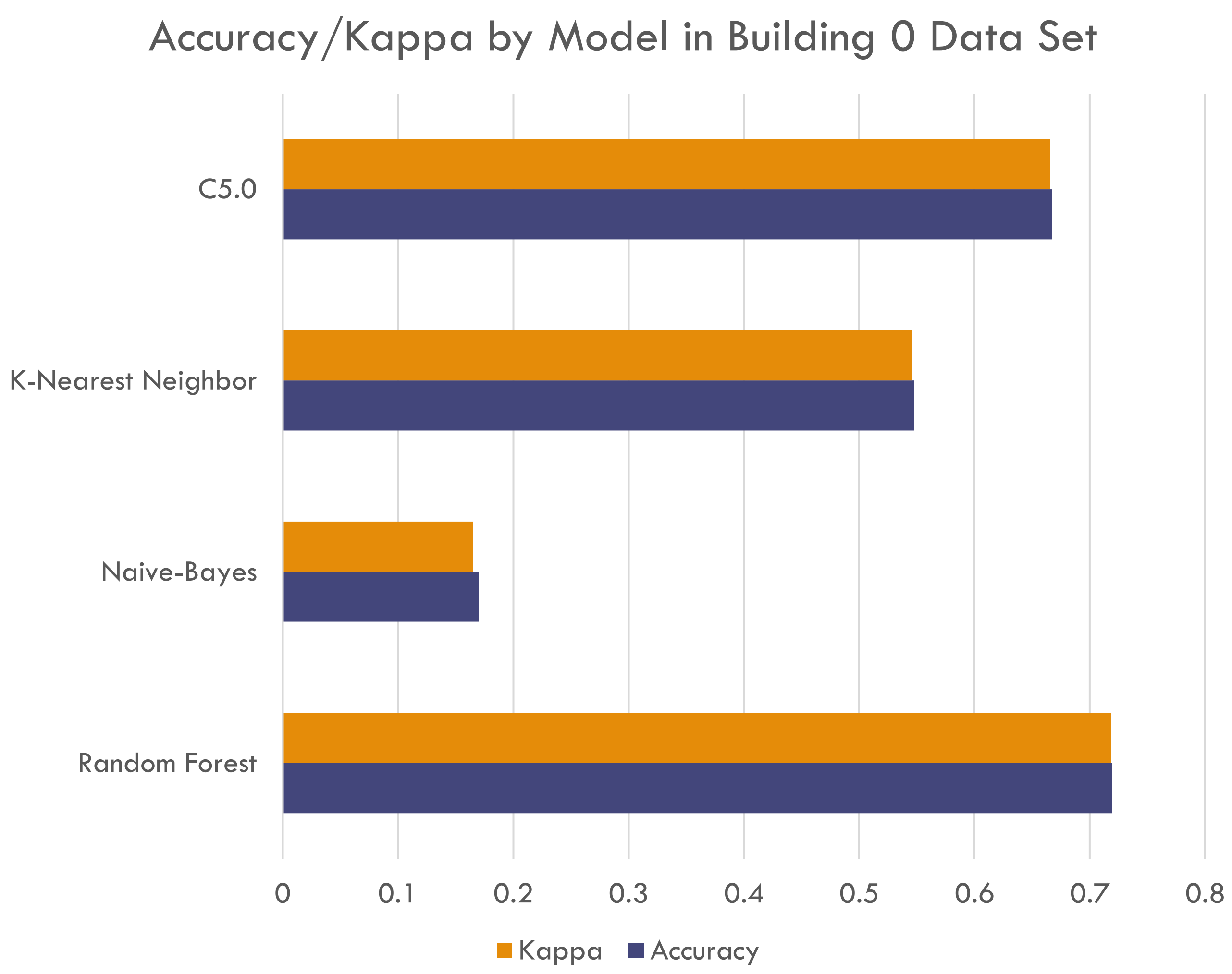
- Accuracy: 0.719431183
  Kappa: 0.7182556
- Our top performing model.



MODEL PERFORMANCE

Classifier	Accuracy	Kappa
Random Forest	**0.7194312**	**0.7182556**
Naive-Bayes	**0.1701336**	**0.1652163**
K-Nearest Neighbor	**0.5476298**	**0.5457635**
C5.0	**0.6670973**	**0.6657256**

MODEL PERFORMANCE



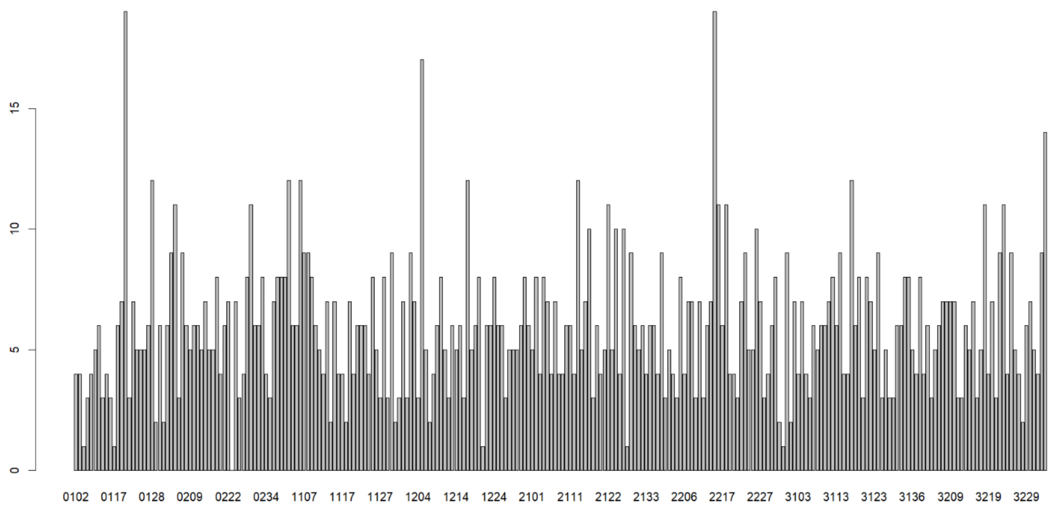
MODEL SELECTION AND VALIDATION

Since the validation data contained null values for SPACEID, we used our test set (30% of the data) to perform predictions.

How well did our model perform with predicted values?

- Accuracy: 0.7636603
- Kappa: 0.7626738

A validation set with complete SPACE ID values may yield a better understanding of model performance.



RANDOM FOREST

RECOMMENDATIONS

SUGGESTIONS FOR FUTURE STUDY

RSSI Impediments

Assess how signal attenuation or interference may reduce the accuracy of observations.

Alternative Methods

Compare fingerprinting to alternative wi-fi locationing methods, such as Time of Flight (ToF) or Angle of Arrival (AoA).

More Observations

More observations (i.e. more WAPs) may increase the accuracy of the models. However, more WAPs will incur more cost.