Wifi Locationing

Kirsten Bjornson

Outline

Overview

The Data

Data Preparation

Modeling

Model Evaluation

Recommendations

Overview

- ➔ Our client is developing a system for large industrial campuses that would determine a person's location within the complex. Because GPS doesn't work well indoors, they want to explore the concept of Wifi Fingerprinting, which uses multiple wifi hotspot signals in order to pinpoint someone's location.
- ➔ Given a database containing wifi signal information and the resulting location, we will evaluate several models and determine which performs best in this scenario.

The Data

- We used data from the UJIIndoorLoc Data Set, which consists of 529 attributes, including:
  - 520 Wireless Access Points (WAPs)
  - Latitude and Longitude coordinates
  - The Building ID, Floor Number, Space ID, and Relative Position within the space
  - User ID, Phone ID, and Timestamp
- Each WAP's corresponding Received Signal Strength Intensity (RSSI) is recorded as a negative integer ranging from -104dBm to 0, meaning that a signal has been detected. A positive value of 100 indicates that no signal has been detected.

Data Preparation

- ➔ To prepare the data for analysis, we first removed any variables that had zero variance. This left us with 465 WAPs to work with.
- ➔ Other irrelevant features were removed, including Latitude and Longitude, User ID, Phone ID, Timestamp, and Relative Position
- ➔ Next, the data was separated by Building number (0, 1, and 2).
- ➔ Then, a unique identifier combining the Building ID, Floor Number, and Space ID was created, called “Location”, and was used as the dependent variable
- ➔ Data was then ready for model testing

Modeling

Model Comparison

Initially, 3 different models were testing on the Building 0 data subset. The Kappa and Accuracy Results are compared in the adjacent table.

Random Forest had the best performance of the 3 models.

Model Performance for Building 0

	Kappa	Accuracy
KNN	0.5480091	0.5499506
SVM	0.5825174	0.5843254
Random Forest	0.7552358	0.756296105

Random Forest

Because Random Forest performed so well with the Building 0 data, it was used on the remaining Building 1 and Building 2 data. Results are as shown.

Random Forest Performance for each Building

	Kappa	Accuracy
Building 0	0.7552358	0.756296105
Building 1	0.849669574	0.8510923
Building 2	0.813676191	0.8144789

Model Evaluation

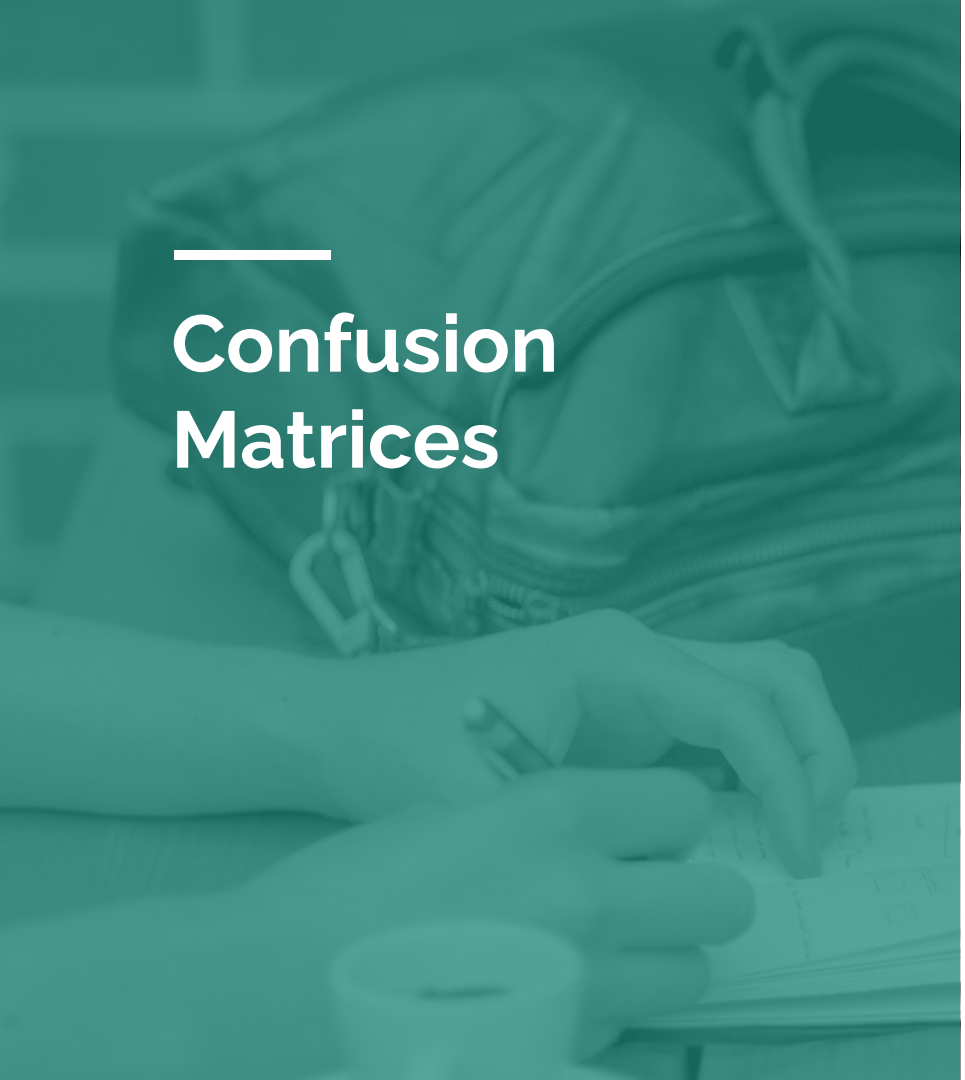
Resampling

Resampling was used to compare the performance across models.

As depicted on the following slide, we see again that the Random Forest Model has the best Accuracy and Kappa metrics of the models tested (for Building 0).

Resampling Results

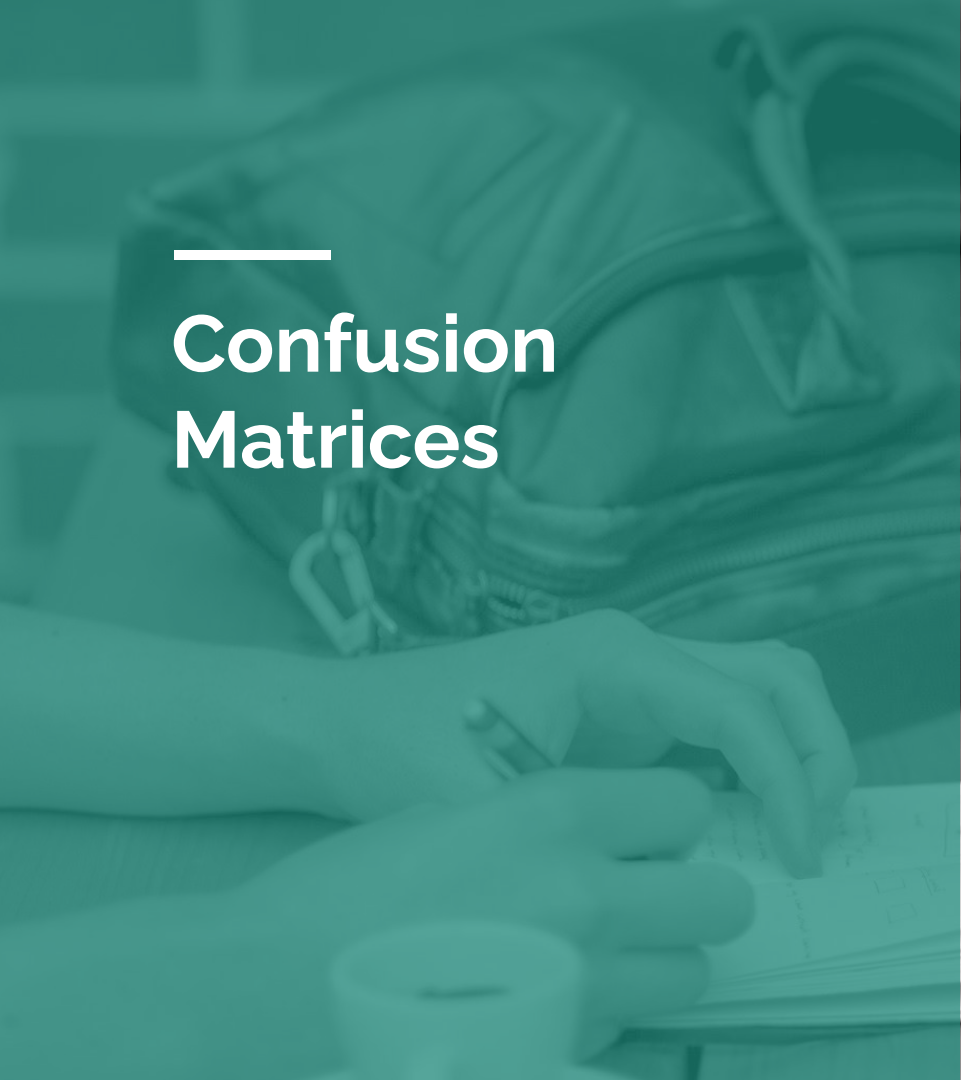
Accuracy						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
KNN	0.5331695	0.5432706	0.5467238	0.5499506	0.5570139	0.57
SVM	0.5379147	0.5725665	0.5794322	0.5843254	0.6031137	0.6181818
RF	0.70694	0.7430621	0.7493796	0.7562961	0.7760136	0.798005
Kappa						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
KNN	0.5311572	0.5413101	0.5447527	0.5480091	0.5550923	0.5681349
SVM	0.5359578	0.5706844	0.5775766	0.5825174	0.6013753	0.6165261
RF	0.7056883	0.7419322	0.7483056	0.7552358	0.7750185	0.797132



Confusion Matrices

A confusion matrix comparing the predicted values to actual values was then used to evaluate the performance of Random Forest for each building.

Due to the sheer number of variables, it's difficult to evaluate the confusion matrix itself. Instead, we will focus on the resulting metrics, depicted on the next slide.



Confusion Matrices

Results:

	Building 0	Building 1	Building 2
Accuracy	0.767	0.8624	0.827
95% CI	(0.7425, 0.7901)	(0.8421, 0.8809)	(0.8109, 0.8423)
No Information Rate	0.008	0.027	0.0109
P-Value [Acc > NIR]	< 2.2e-16	< 2.2e-16	< 2.2e-16
Kappa	0.766	0.8611	0.8263

The results of the confusion matrix show that the predictions made for the test set for each building are quite good, with Building 1 having the best overall results for Accuracy and Kappa.

Recommendations

Of the algorithms tested, Random Forest performed the best by far when looking at the Kappa and Accuracy values. Evaluation of the models with Resampling back up the initial Random Forest Results, as well as using a Confusion Matrix to compare the models' predicted values to actual values.

Models performed well when combining Building ID, Floor, and Space ID into one variable. Separating the data by building was helpful for initial analysis. Further testing on combined building data should be performed and compared.

Recommendations

Other modeling methods such as Ensemble Learning can be tested.

While the Random Forest model produced rather good results, keep in mind that RSSI values can fluctuate due to interference and other factors, making it difficult to pinpoint an exact location.

The system should be updated and calibrated frequently in order to keep up with changes in the building’s environment.

Questions?