



# C3T3 Indoor Wifi Locationing

Presented by Don Bice

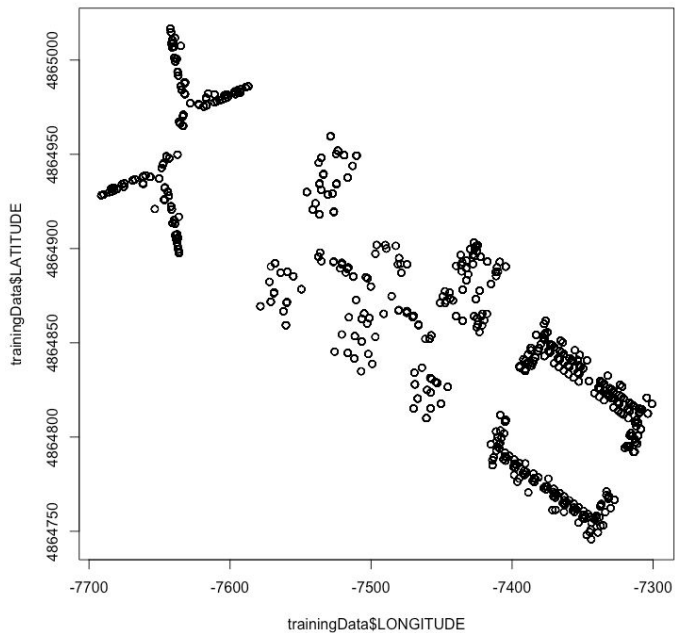
Aug 3rd, 2018

**IOT**  
**ANALYTICS**

# Indoor Location Methods Based on WiFi Fingerprinting



<http://archive.ics.uci.edu/ml/datasets/UJIIndoorLoc>



## UJIIndoorLoc Data

1

21,048 total observations split among 3 buildings w/ 4-5 floors.

2

529 total features. 520 Wireless Access Points (WAPs) as predictors.

3

2 Datasets from original study: trainingData & validationData.

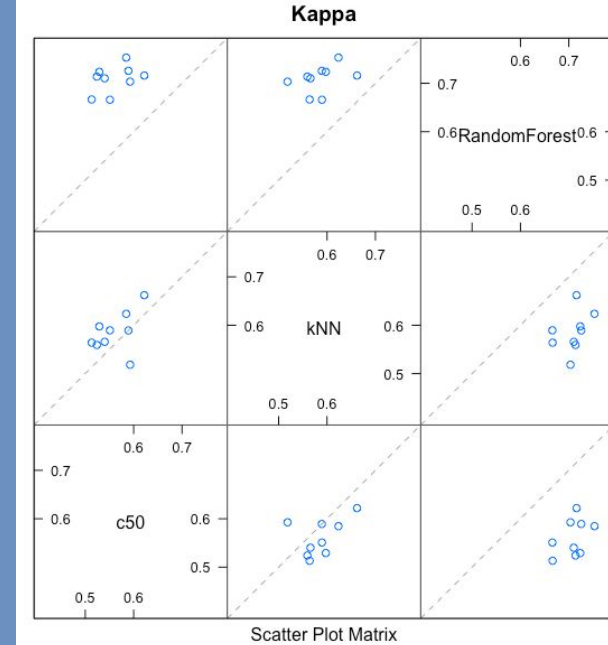
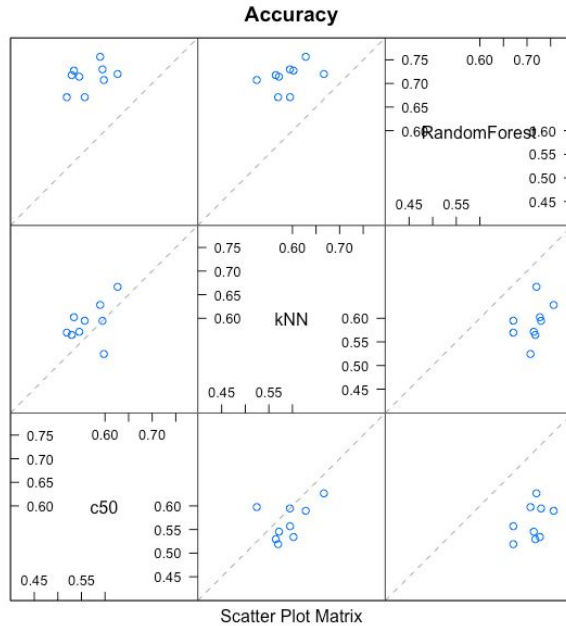


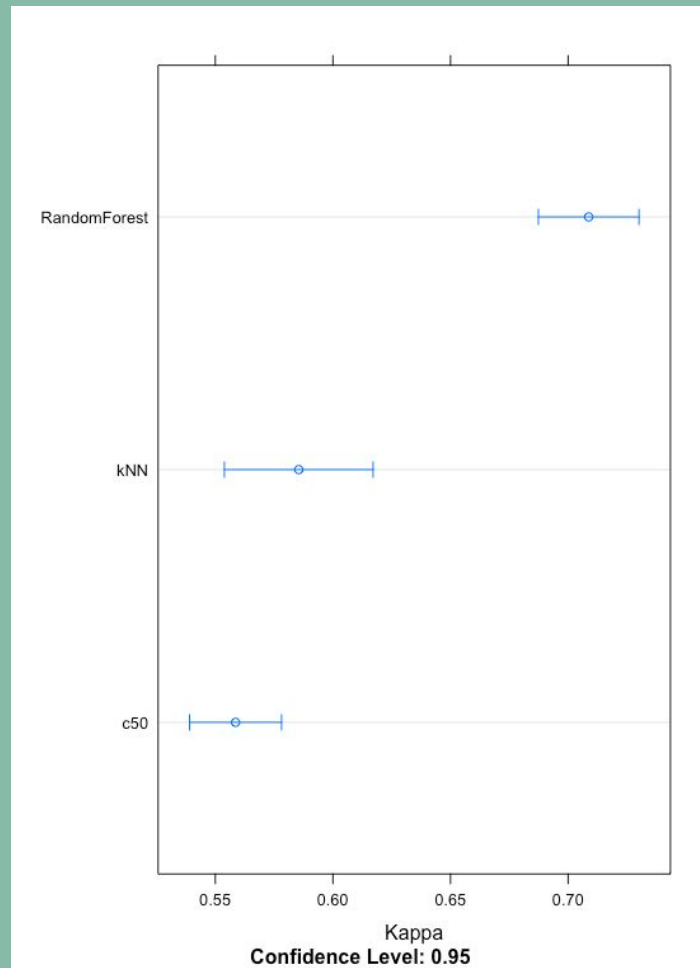
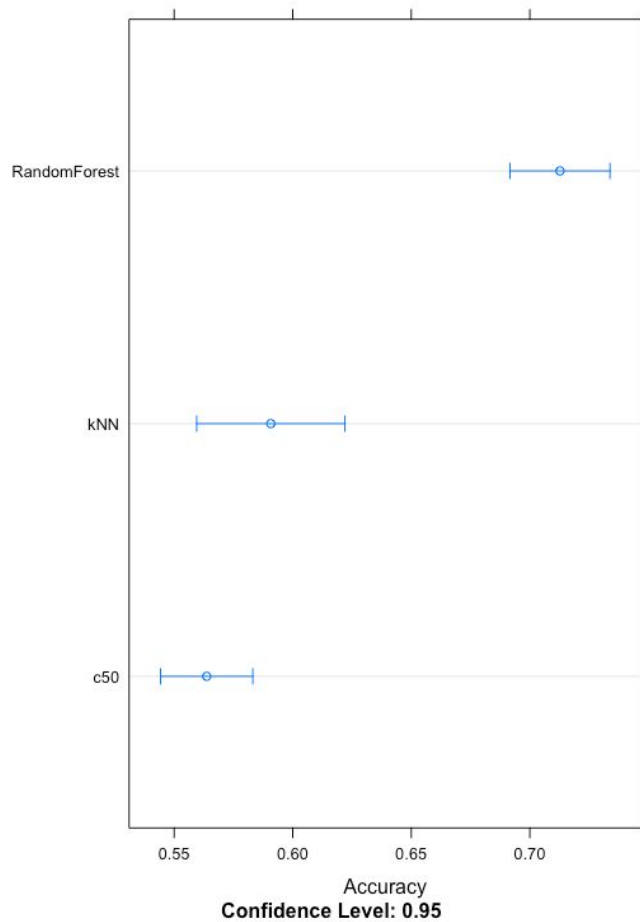
# Classification Methodology

Evaluate multiple machine learning models for best accuracy.

- ❑ Used only “trainingData.csv” from the source data since ~20k observations of 529 variables.
  - ❑ Only 312 WAPs in common with validationData.csv
  - ❑ ValidationData had a number of missing attributes.
- ❑ Combined Floor & SpaceID features into composite variable as target to be predicted.
  - ❑ Experimented with other composite variables.
- ❑ Subsetted observations by 3 buildings and sampled 20% of building 1 due to lower density.
- ❑ Removed features (longitude, latitude, building ID, relative position, timestamp, phone ID, and user ID).
- ❑ Trained and tested using C50, RandomForest, kNN & SVM algorithms to determine best classification model.

# C50, RandomForest & kNN





# RandomForest Was the Best Performer



30 resamples

ACCURACY	Min	Mean	Max
c50	0.440	0.564	0.683
kNN	0.524	0.591	0.667
RandomForest	<b>0.671</b>	<b>0.713</b>	<b>0.756</b>

KAPPA	Min	Mean	Max
c50	0.434	0.559	0.678
kNN	0.518	0.586	0.662
RandomForest	<b>0.666</b>	<b>0.709</b>	<b>0.753</b>

# Trained Model Accurate with New Data

RandomForest



	Accuracy	Kappa
Training	0.713	0.709
Prediction	<b>0.777</b>	<b>0.774</b>

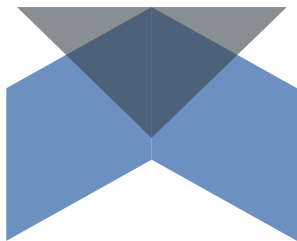


# Conclusions

A lot of iterations.

- ❑ Best model for predicting Floor\_SpaceID composite variable was RandomForest with mtry = 64.
- ❑ Size of data set and dimensionality resulted in preliminary attempts that had very high computational costs and poor accuracy.
- ❑ Subsetting by building and sampling 20% of data offered good accuracy with reasonable computation times as a trade-off.
- ❑ The SpaceID identifying an office or classroom in the data is unconventional and not unique (multiple spaces on multiple floors in multiple buildings with same SpaceID).
- ❑ WAPs are unevenly dispersed geographically and in some areas likely too dense to allow for accurate triangulation/positioning. Poor communications topology.





| IOT ANALYTICS