Evaluating techniques for Wifi-locationing indoors

Team members at IOT Analytics:
Ari | Madoka | Violeta

Proposals of best analysis:

The team has worked individually to study machine learning algorithms to create models that most accurately predict locationing indoors.

Seeing the possibility to continue on the project by incorporating this model to a smartphone app, we decided to form a Team and present to you the different approaches.

With this we hope to arrive to the decision on how is best to proceed and work together in delivering a high-quality result to our client.

The guideline

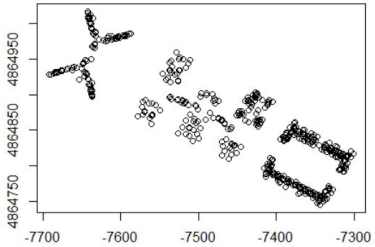
- Exploration of the dataset
- Pre-processing
- Exploring/Graphing
- Recommendations

- **Exploration of the dataset**
- Pre-processing
- Exploring/Graphing
- Graphing predictions
- Recommendations

The guideline

Exploration of the dataset

Wifi-fingerprinting with 2.4Ghz



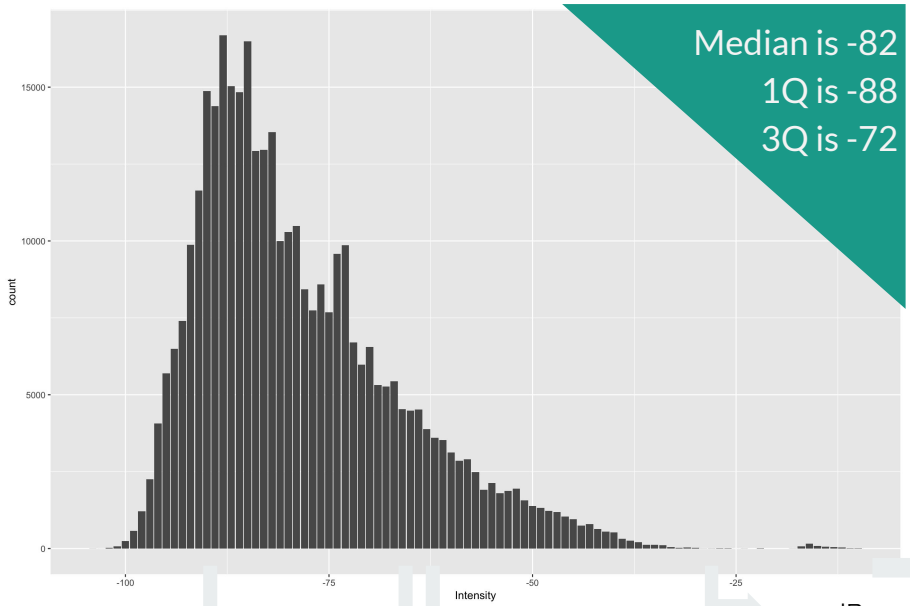
Min timestamp : May 30th, 2013
Max timestamp: June 20th, 2013

529 attributes, 520 are WAPS

Universitat Jaume I, Valencia

Exploration of the dataset

Data composition:



Intensity of 100 is 96% present in WAPS

-30 dBm = Amazing

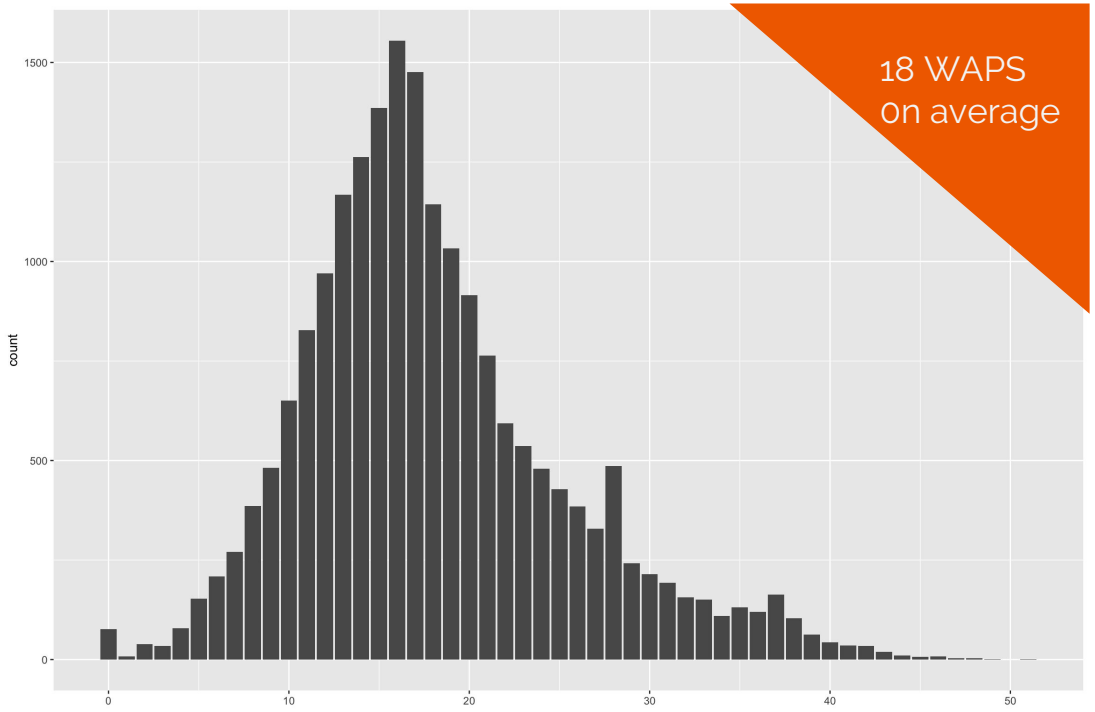
-67 dBm = very good

-70 dBm = Okay

-90 dBm = Unusable

Exploration of the dataset

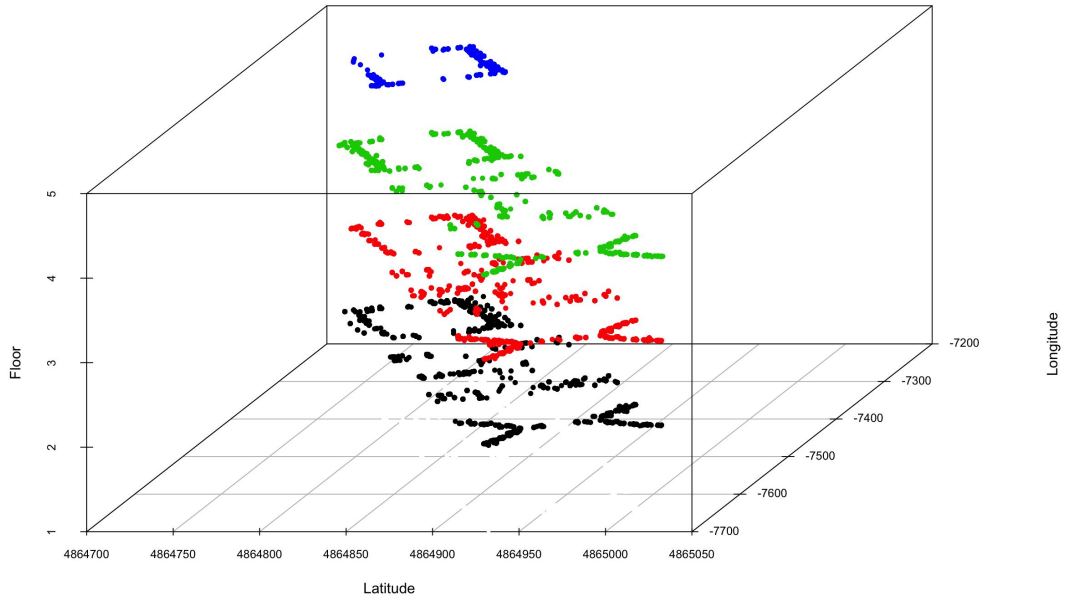
Visualizing active WAPS:



520 waps with recorded information

Exploration of the dataset

Points of measurement per building:



- Exploration of the dataset
- **Pre-processing**
- Exploring models
- Graphing predictions
- Recommendations

The guideline

Pre-processing

Same approaches:

1. Turning BuildingID and FloorID into factors

2. Keep relevant attributes:
BuildingID, FloorID, Latitude, Longitude

3. Ignoring WAPS with close to near zero variance which eliminates 50WAPS

4. Converting +100 intensity to -110 intensity for normalization purposes

Differences:

1. Combining training and validation datasets for joint measurement. Then splitting.

2. Intensity values normalized with logarithmic scale.
For Lat / Long normalization for dataframe.
For Building, normalization per row.

3. Taking advantage of high accuracy in Building Prediction, subsetting respectively.

4. Sample sizes between 3000 and 8000

Pre-processing

Exploring Normalization:

Given Training Data

LONGITUDE	LATITUDE	FLOOR	BUILDINGID
Min. :-7691	Min. :4864746	0:4369	0:5249
1st Qu.:-7595	1st Qu.:4864821	1:5002	1:5196
Median :-7423	Median :4864852	2:4416	2:9492
Mean :-7464	Mean :4864871	3:5048	
3rd Qu.:-7359	3rd Qu.:4864930	4:1102	
Max. :-7301	Max. :4865017		

Given Validation Data:

LONGITUDE	LATITUDE	FLOOR	BUILDINGID
Min. :-7696	Min. :4864748	0:132	0:536
1st Qu.:-7637	1st Qu.:4864843	1:462	1:307
Median :-7560	Median :4864915	2:306	2:268
Mean :-7529	Mean :4864902	3:172	
3rd Qu.:-7421	3rd Qu.:4864967	4: 39	
Max. :-7300	Max. :4865017		

Arguments:

1. Normalizing after combining

2. Non-significant variance + large observations in Training Data 19937 vs 1111

- Exploration of the dataset
- Pre-processing
- **Exploring models**
- Graphing predictions
- Recommendations

The guideline

The goal:

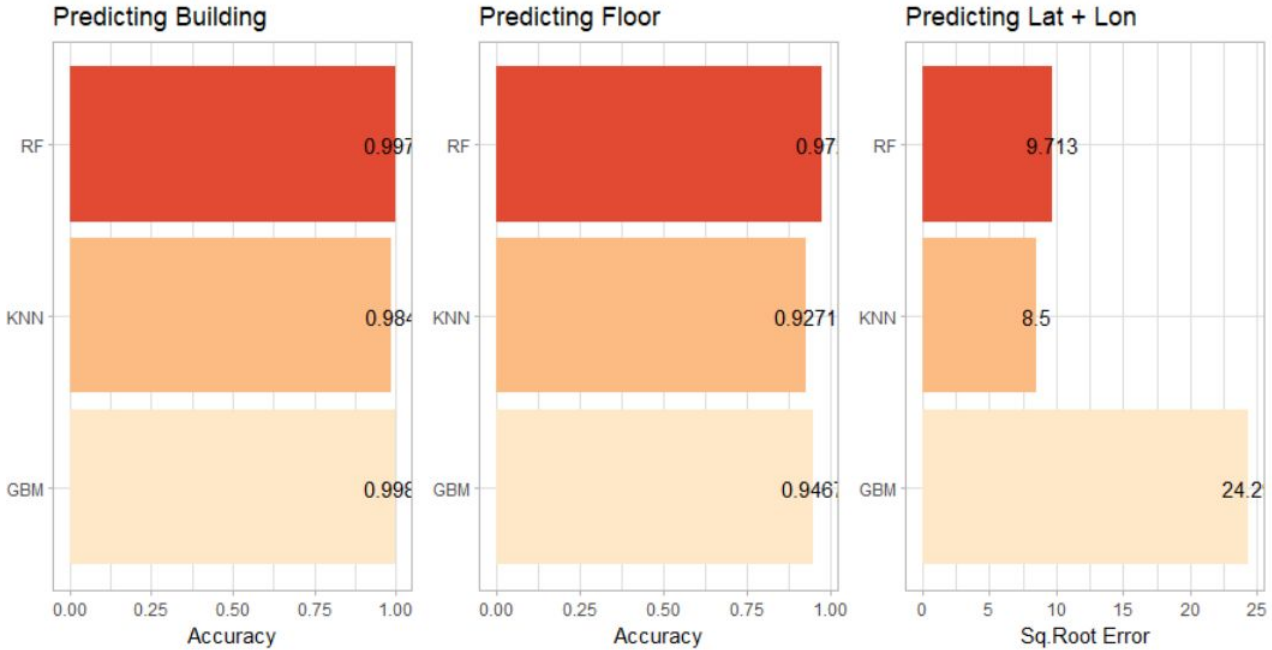
Locate a person inside a Building based on her Longitude, Latitude, and Floor location

Methods:

Classification and Regression Machine Learning Techniques: KNN, RF, GBM

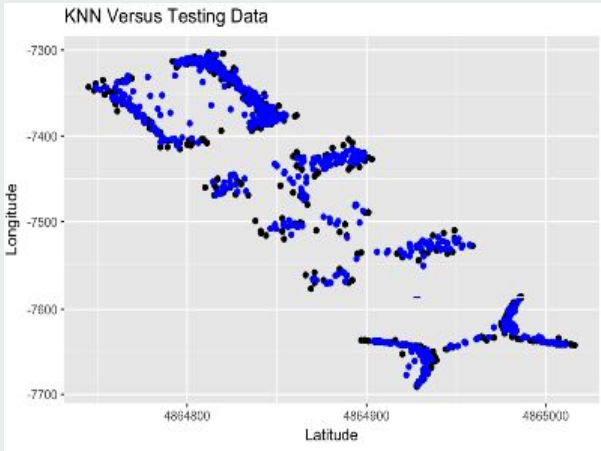
Exploring models:

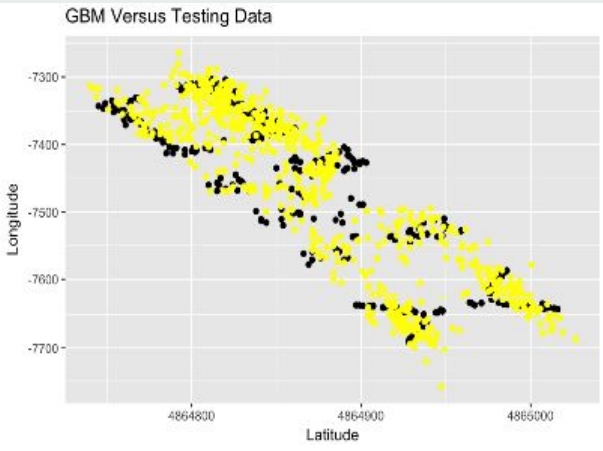
APPROACH 1: Splitting dataset containing four attributes (BuildingID, FloorID, Latitude, Longitude)

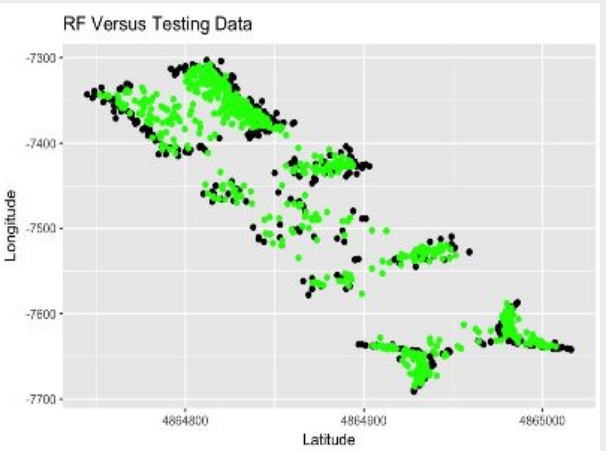


Graphing Predictions

APPROACH 1: VISUALISING RESULTS

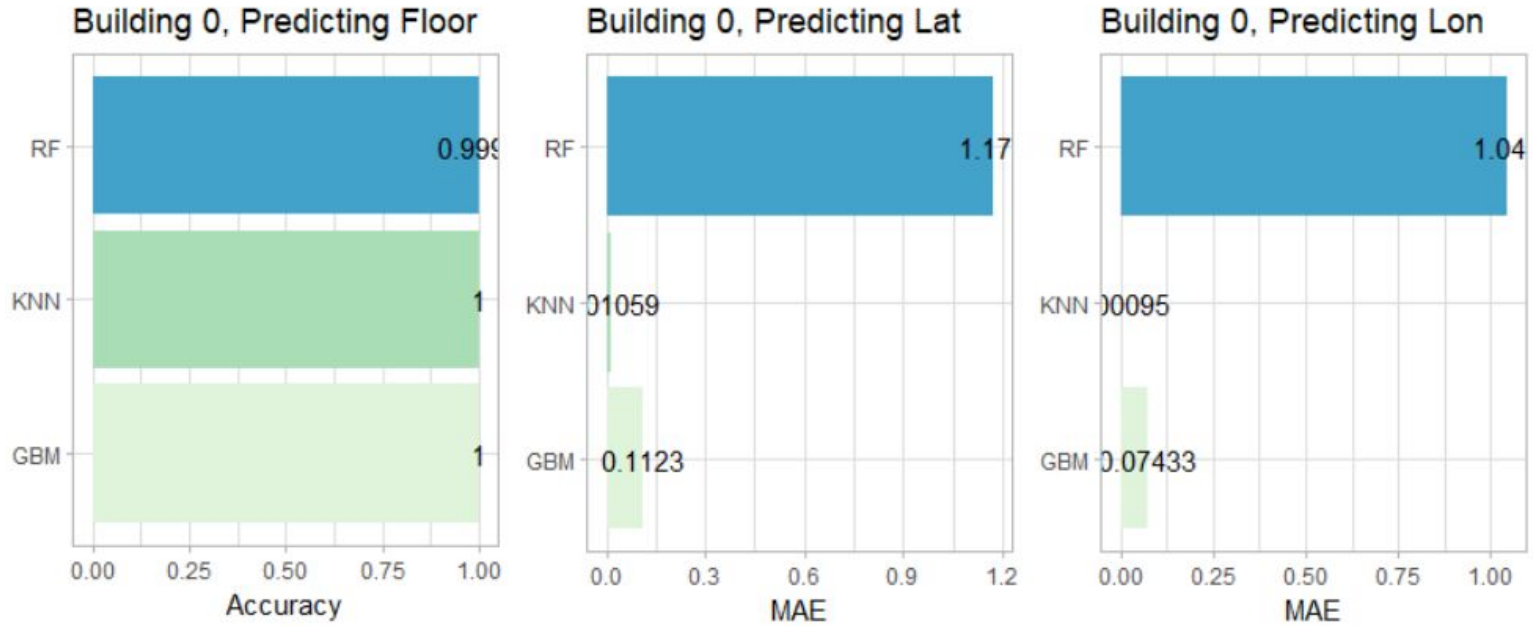






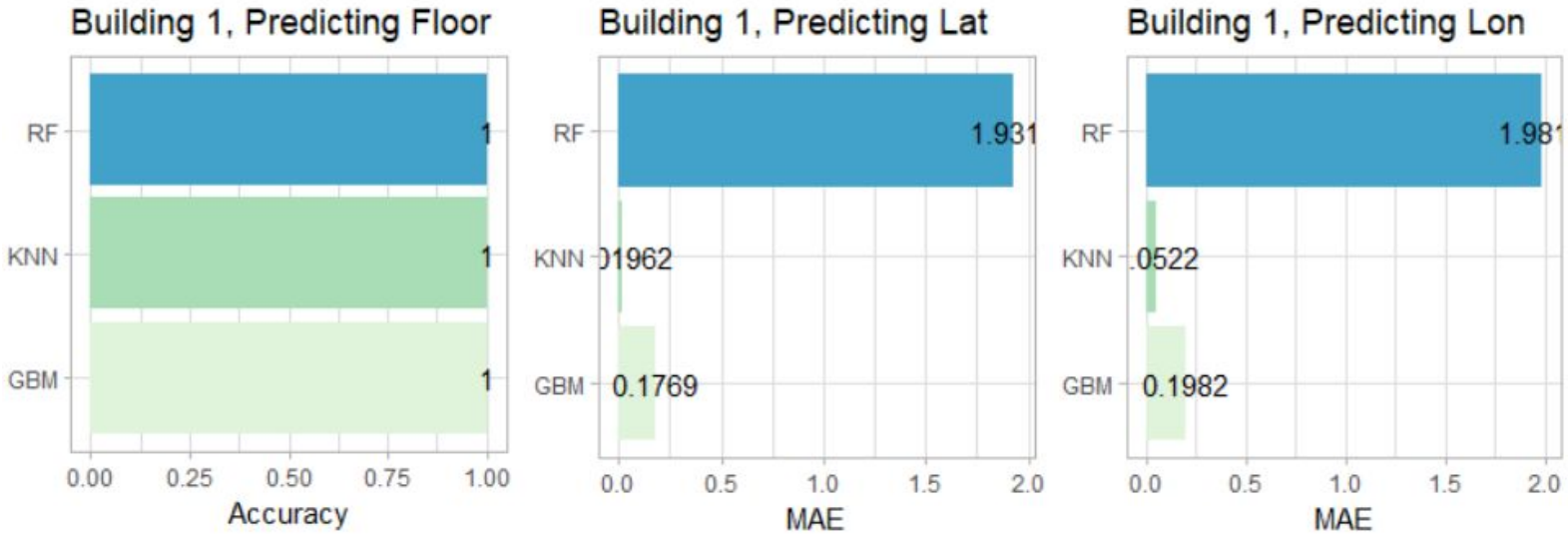
Exploring models

APPROACH 2:. Predicting BuildingID + subsetting according to floor, latitude, longitude



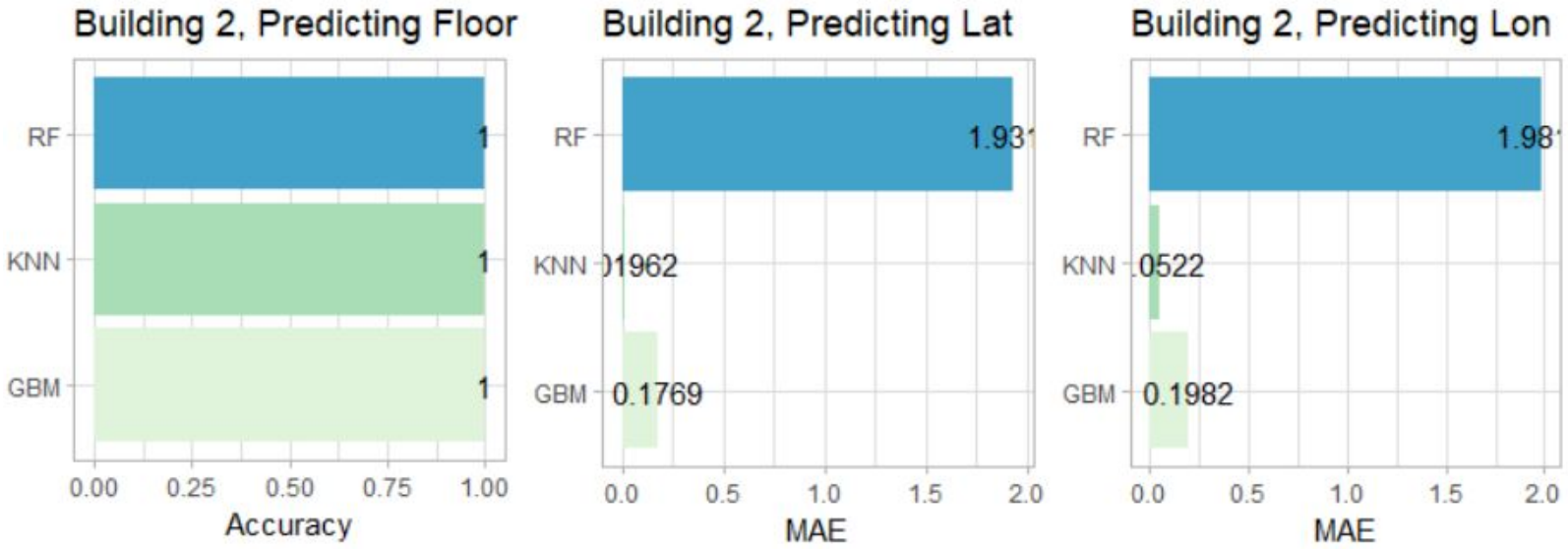
Exploring models

APPROACH 2:. Predicting BuildingID + subsetting according to floor, latitude, longitude



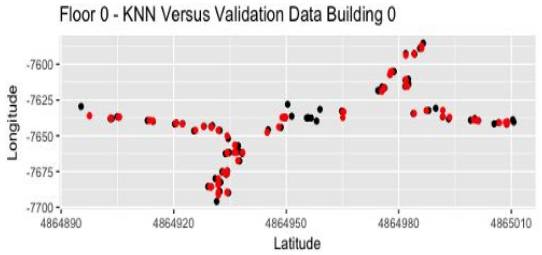
Exploring models

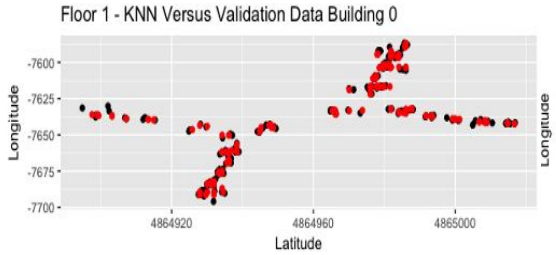
APPROACH 2:. Predicting BuildingID + subsetting according to floor, latitude, longitude

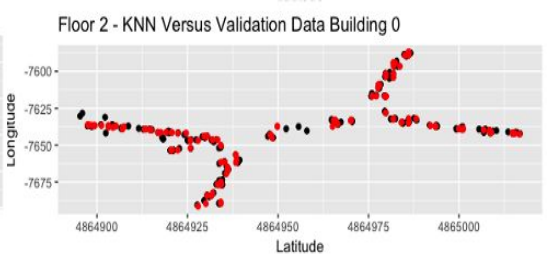


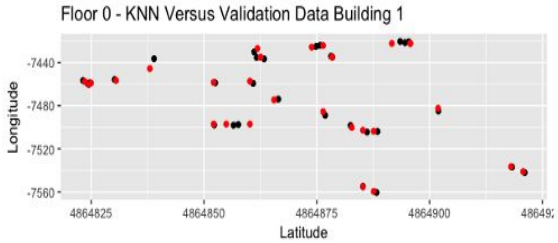
Graphing Predictions

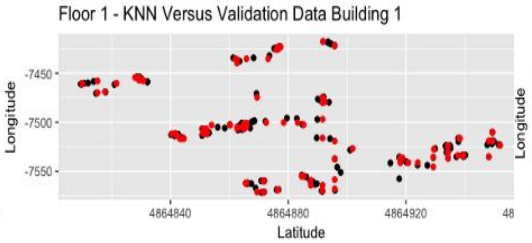
APPROACH 2: VISUALISING RESULTS

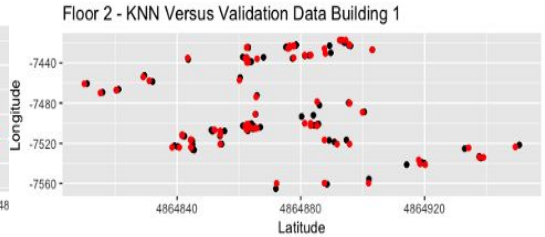






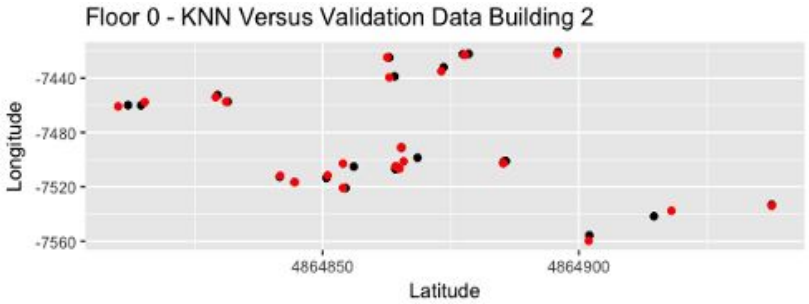


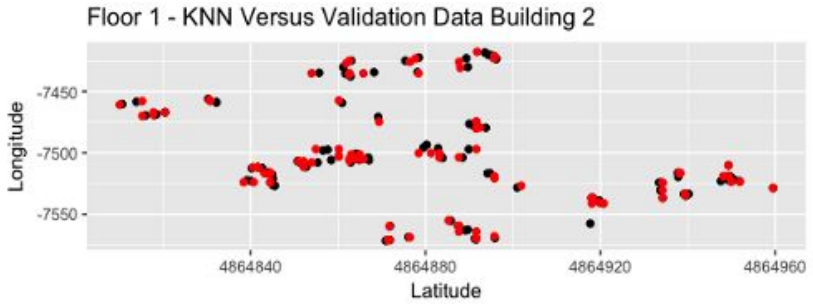


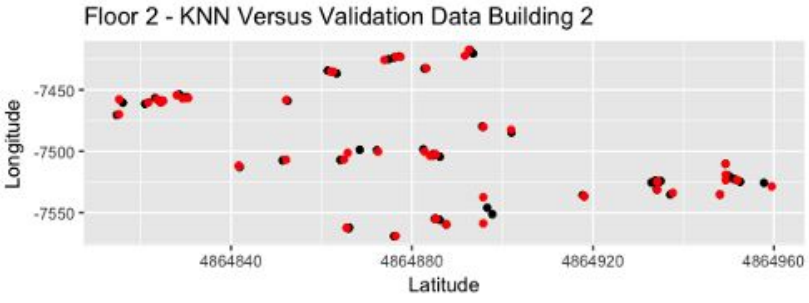


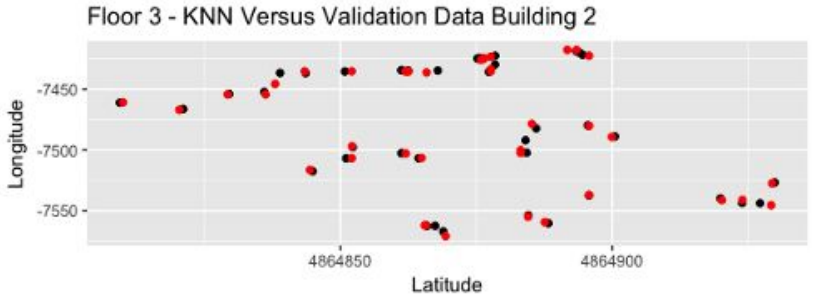
Graphing Predictions

APPROACH 2: VISUALISING RESULTS









- Exploration of the dataset
- Pre-processing
- Exploring/Graphing
- **Recommendations**

The guideline

Recommendations

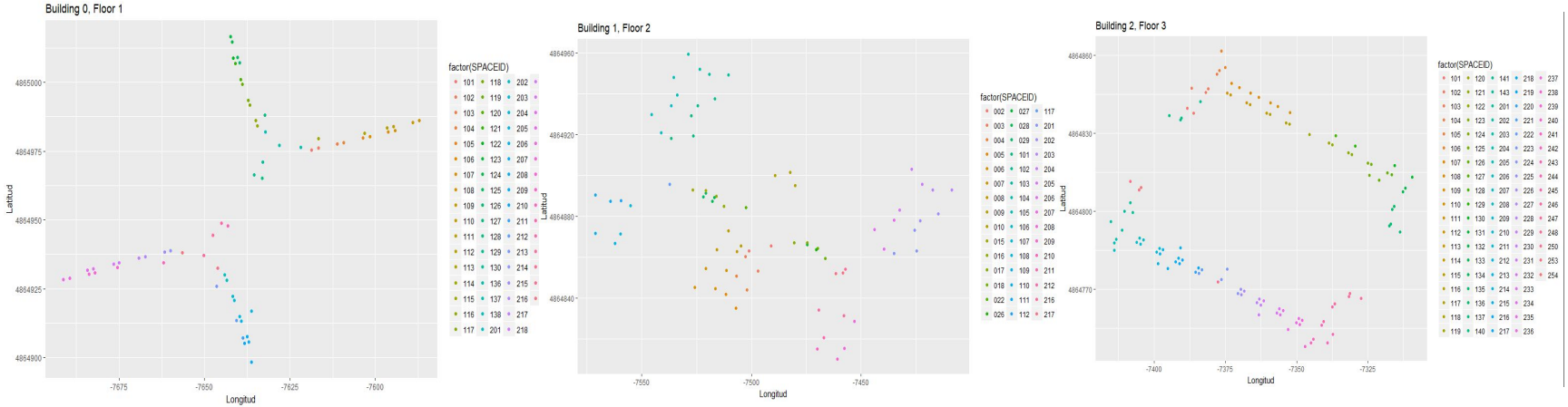
1. Combination of techniques

2. Wise sampling

3.Studying ‘Dark data’ given sources

http://www.sbrt.org.br/sbrt2016/anais/ST13/1570275280.pdf
http://journals.sagepub.com/doi/full/10.1155/2015/429104

Mapping Space ID per Building and Floor:



Thank you for listening,

Let’s discuss!