IoT Analytics

WI-FI FINGERPRINTING

Vera Rykalina, July 2020

Outline

- *Challenge*
- *Project Objectives*
- *Use Cases*
- *Data Overview*
- *Datasets*
- *Methodology*
- *Perfomance of Models*
- *Error Estimation*
- *Recommendations*
- *Always Ways To Improve*

Challange

- As opposed to outdoor positioning based on GPS satellites, indoor localization with a mobile phone is not a trivial task because of either weak or no access to satellite signals.
- Dynamic research to tackle this problem revolves around the use of signals from Wireless Access Points (WAPs) in a building.

Project Objectives

- Assess feasibility of Wi-Fi Fingerprinting leverage to determine a person's location in indoor spaces by evaluating multiple machine learning models.

- Apply Received Signal Strength Indicators (RSSI) values recorded from multiple Wi-Fi hotspots within the building as input data for modelling.

- Compare classification (Accuracy & Kappa) and regression (RMSE & MAE) algorithms and select the best performance.

Use Cases

- Shopping Centers
- University Campuses
- Public Airports
- …

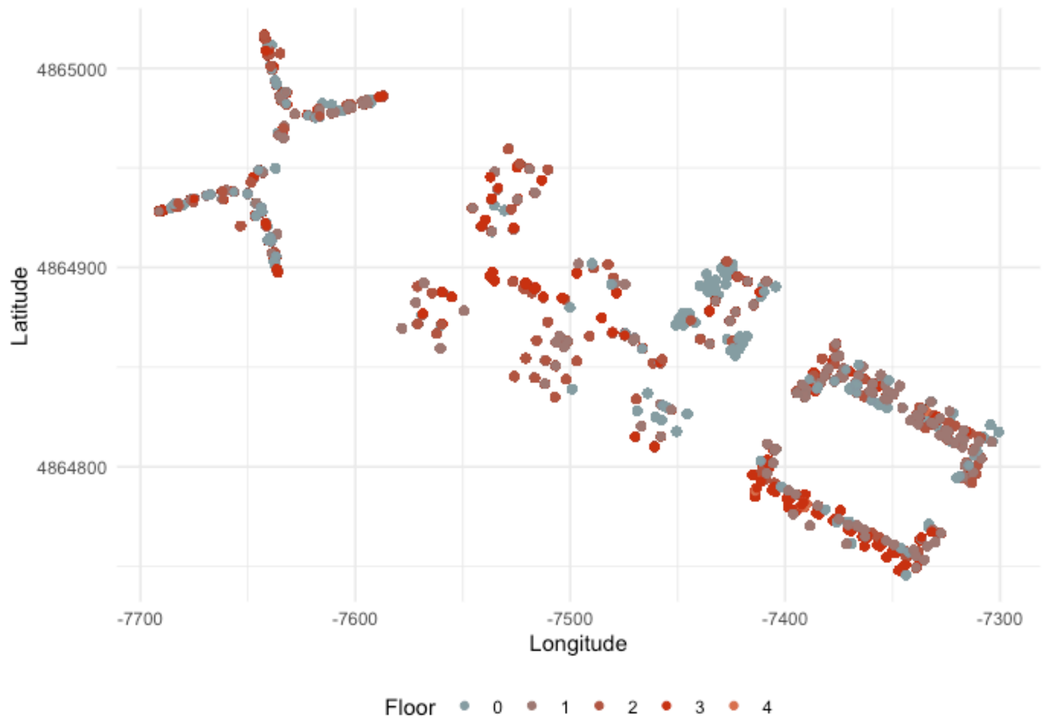
- Door-To-Door Public Transport Navigation
- Blind and Visially Imapaired Community
- Home Automation
- …

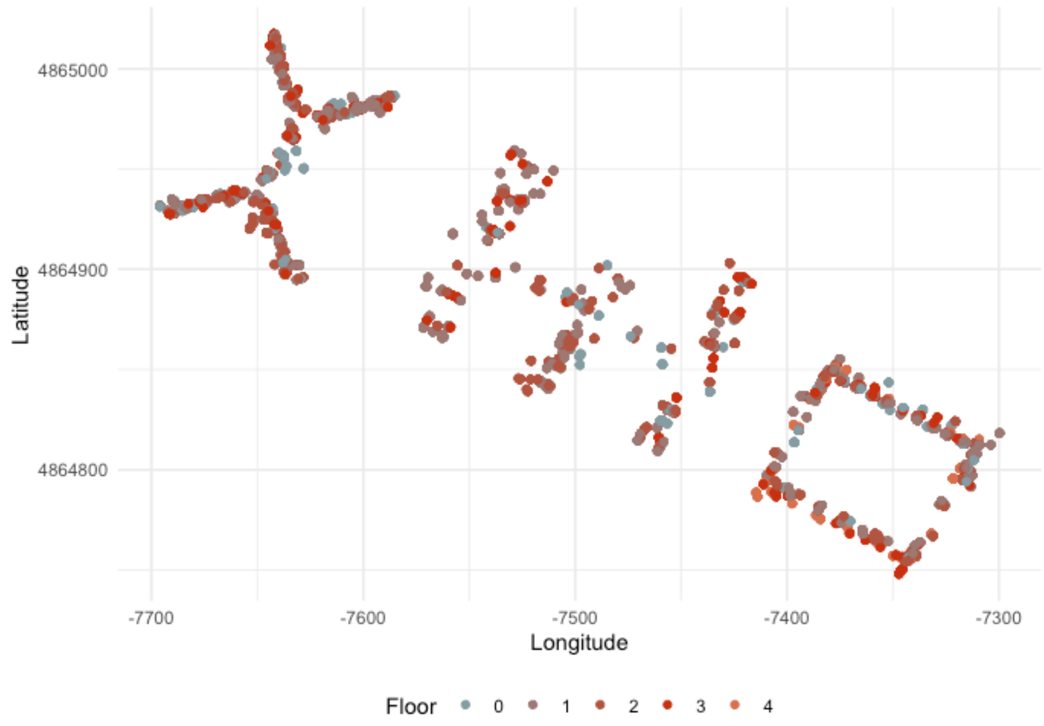
Project Features

Data Overview

- Area of 1087703 m^2
- 933 reference points
- 21049 records: 19938 (training) and 1111 (validation)
- Total WAPs 520
- Data collection: 20 useres and 25 mobile devices
- 4 Months difference between training and validation data collection

Datasets





1- TRAIN (TEST) a large database of Wi-fi fingerprints for a multibuilding industrial campus with a location (building, floor, location ID) associated with each fingerprint.

2- VALIDATION containing the expected results already; allows to check the model performance.

Methodology

1 PRE-PROCESSING

- Missing values treatment = -105
- Data classes conversion = hablar:: convert()
- Zero variance exploration & treatment =caret::nearZeroVar()
- Duplicates = base::unique()
- Descriptive analysis (long format df)
  - Visualization Tool: ggplo2
- Outliers treatment (remove -30 & -90 dBm)

2 FEATURE SELECTION & ENGINEERING

- Combining attributes (BuildingFloor)
- Creating Principal Components (PCA)

3 PRE-MODELLING

- Training dataset partition (75/25)
- Cross-validation

4 MODELLING

- Caret (RF, KNN, C50, XGB, GBM)
  - Applying PCA
  - Tuning tuneGrid()

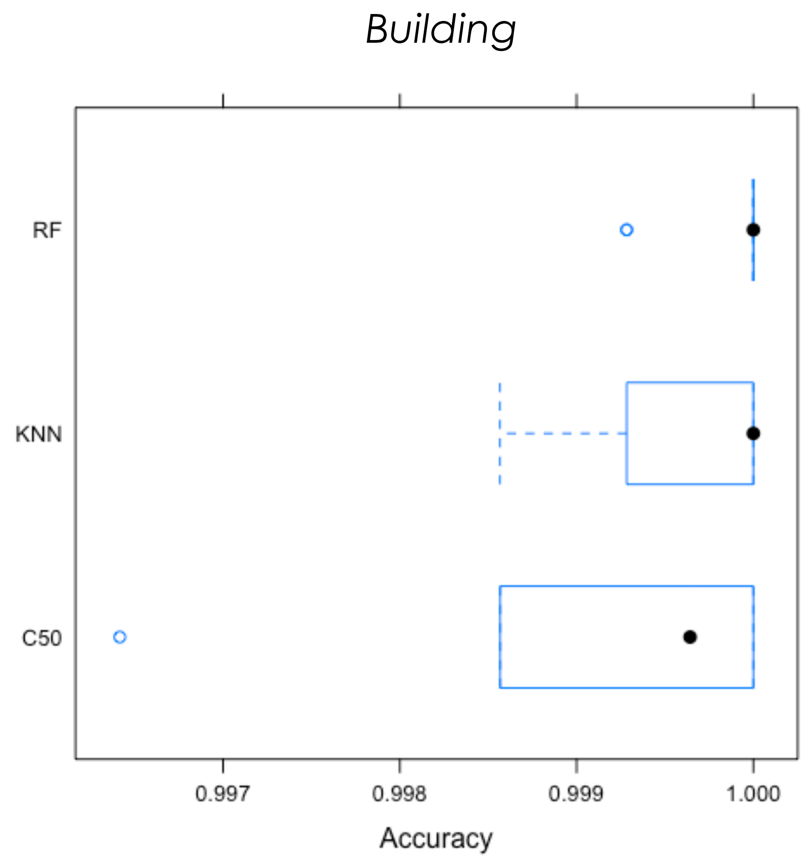
5 ACCURACY AND CONFIDENCE ANALYSIS

- Visualization Tool: ggplo2

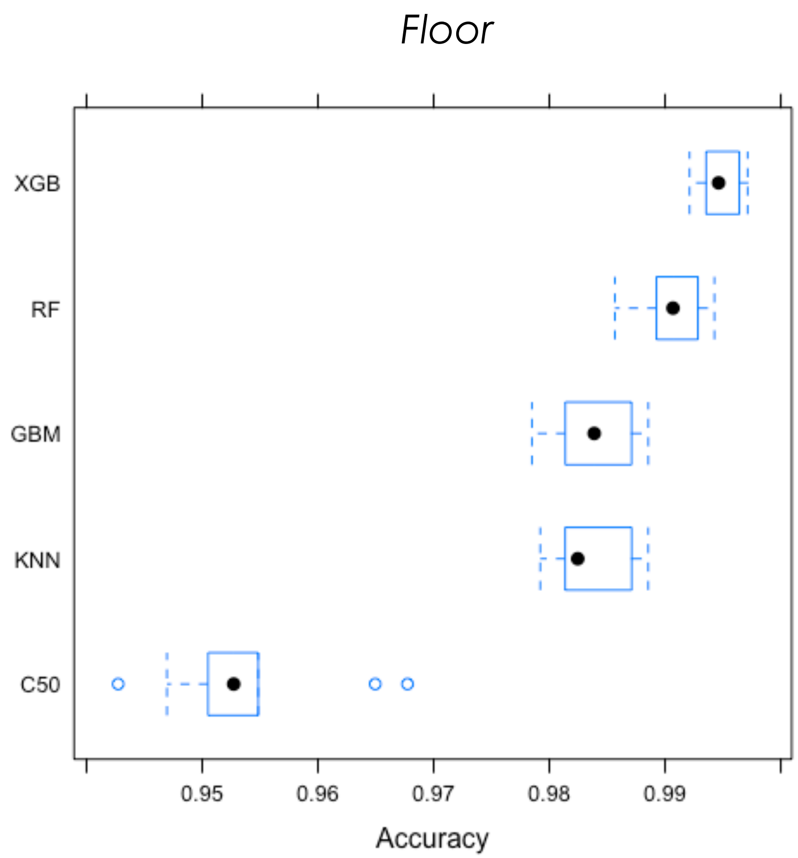
6 PREDICTION

- Validation dataset

Classification

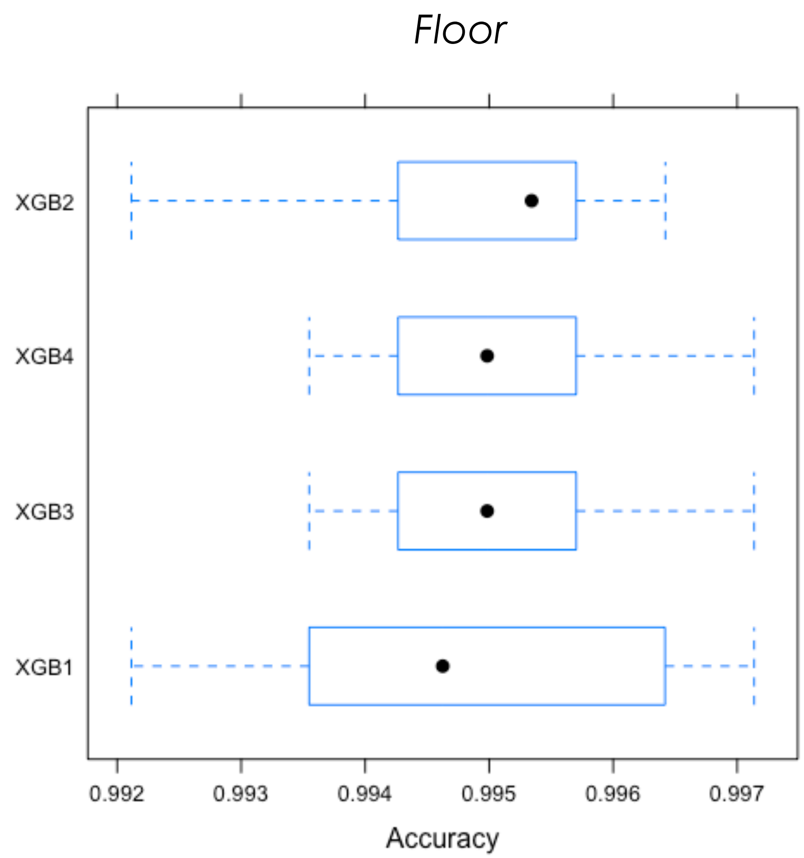


Classification

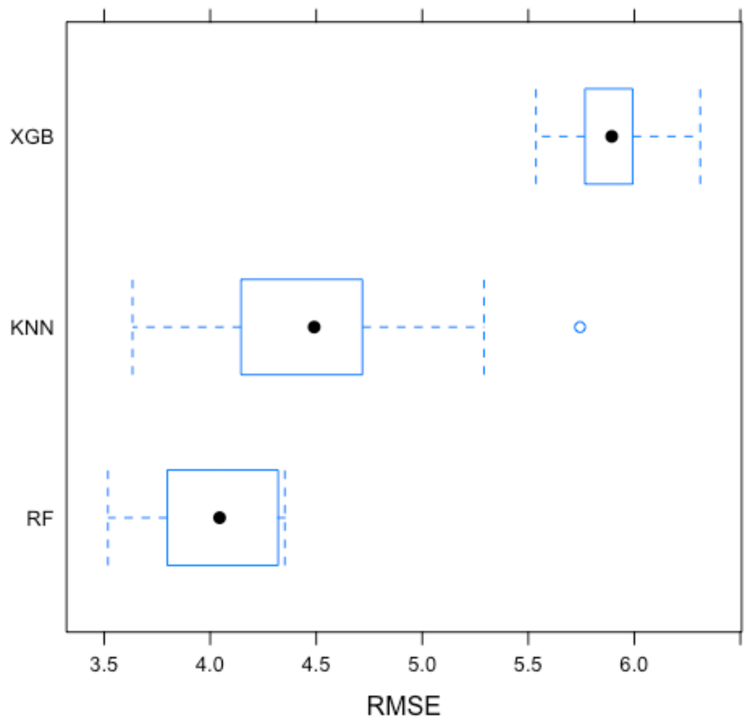


Classification

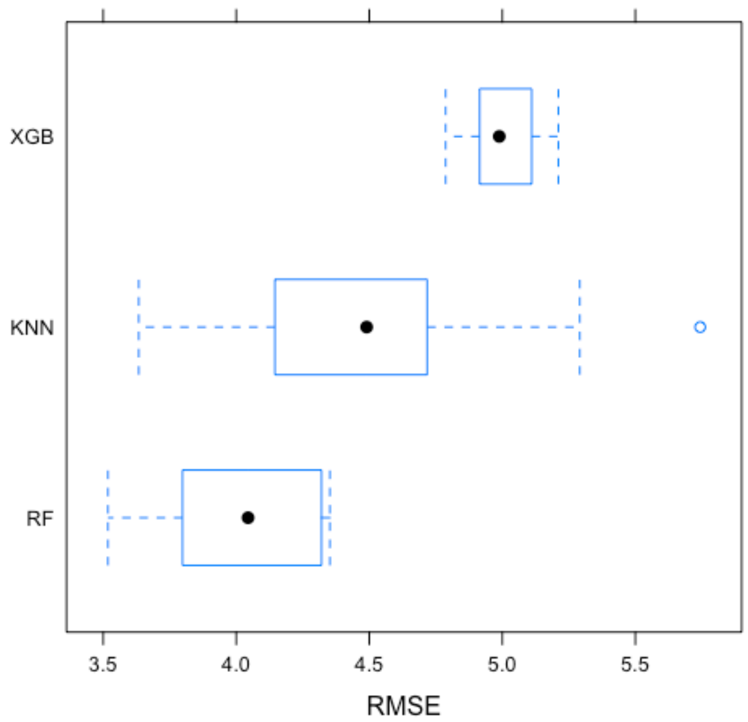
Model Tuning



Regression



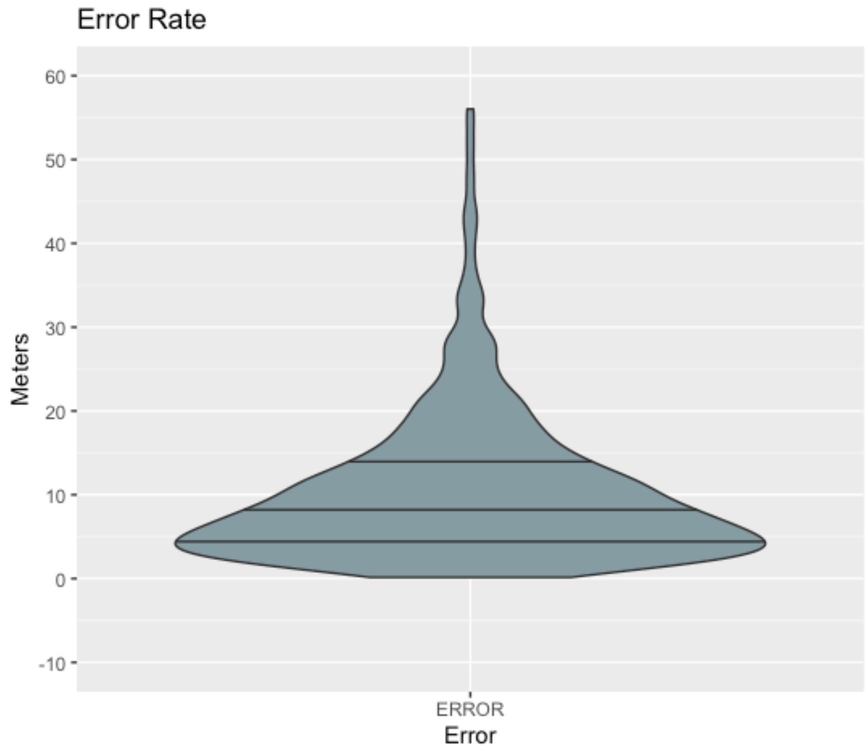
Longitude



Latitude

Error Estimation

25%	50%	75%	100%
4.3001737	7.9456736	13.7360994	132.6573180

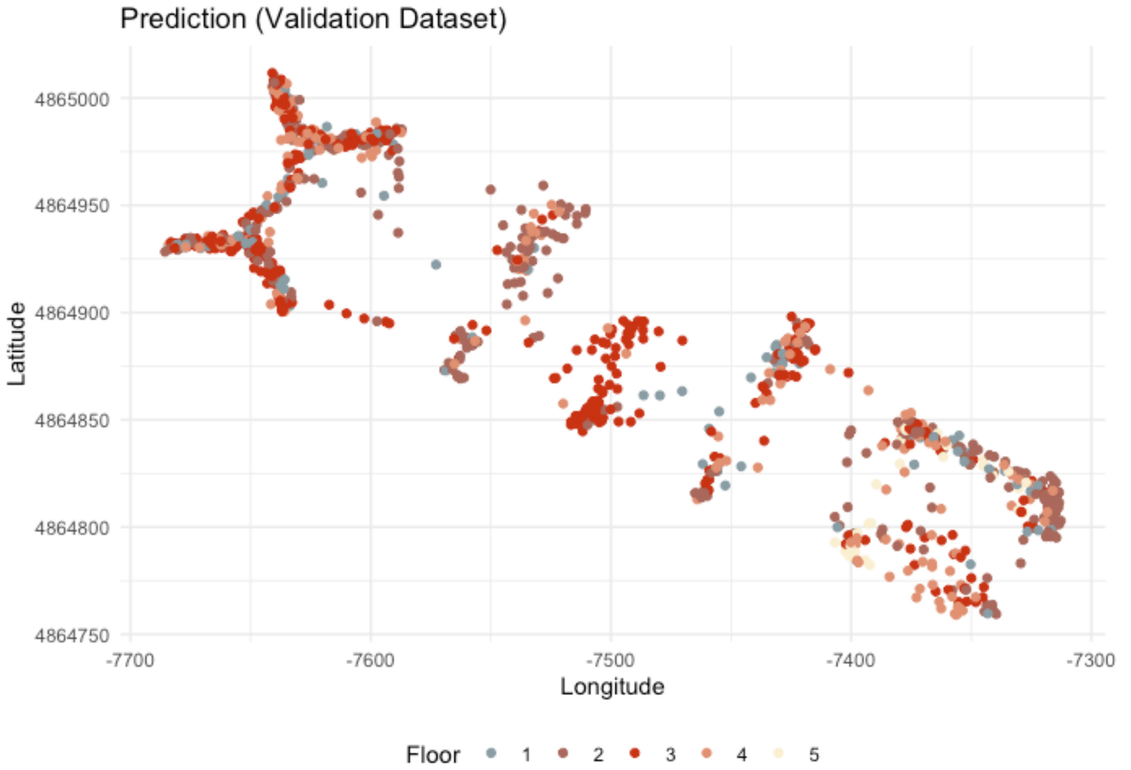


Recommendations

- Solid WAP locations
- Actualization of WAP changes: installation of new and remove of old
- Location of WAPs at buiding cornes (training dataset)
- Time series data as approch (input: last position, speed limit constraint)
- Reasonable accuracy vs computing power
- Disctribution of WAPs can be imprived (better coverage)

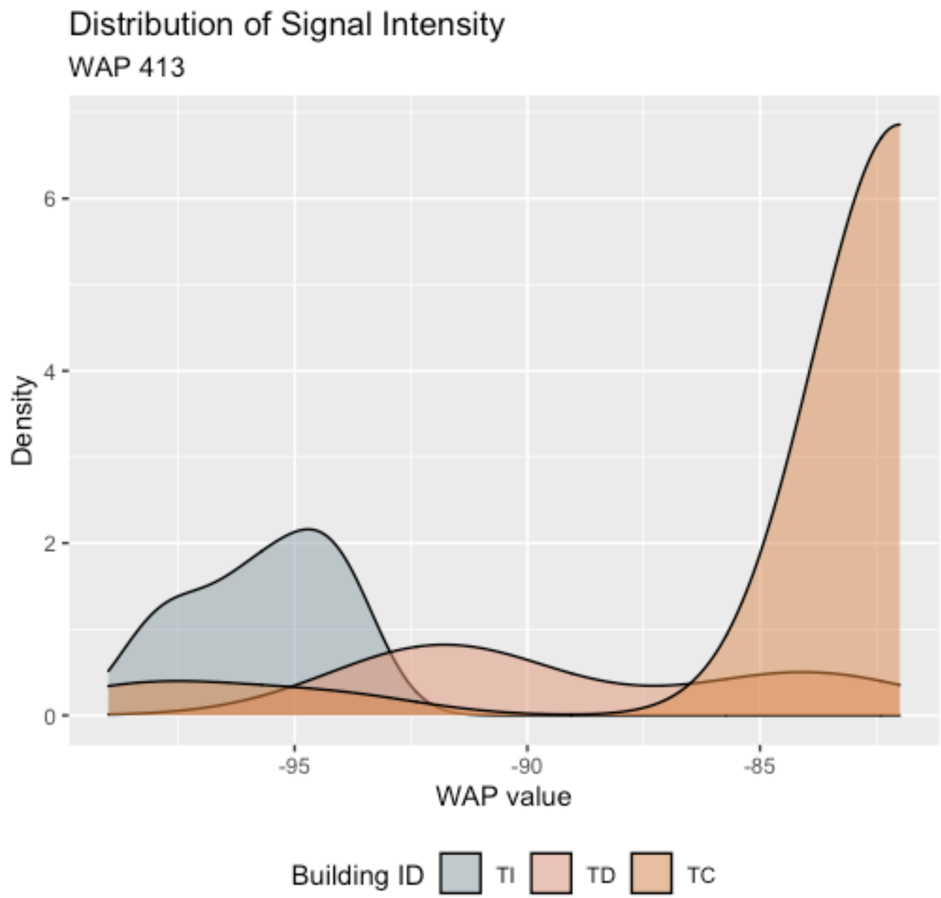
Always Ways To Improve

- Synchronization of training and validation datasets? (timestamps)
- Building-Floor feature engineering
- Trying out WAPs vs PCAs
- Explore outliers in predicted locations (plotly)



Always Ways To Improve

- Smart resolving of signal overlaps → Lon/Lat check (figure)
- Leverage of non-caret algorithms
- Normalization (other methods)



Thank you!