Wifi Location

Oriol Ordi-Fabi Reyes

Introduction

Aim

- The goal of this project is to investigate the feasibility of using "Wifi fingerprinting" to determine a person's location in indoor spaces
-

Data

- Data base from the Universitat Jaume I which contains 19936 observations on the training set and 1111 on the validation set.
- There are 529 attributes of wich 520 belongs to the Wireless Access Points (WAPs).
- The target variables for the predictions are : Building, Floor, Latitude and Longitude.

Table of Contents

1 Data Exploration

1. Analysis of the Distributions
2. Data Points Location

2 Data Pre-procesing

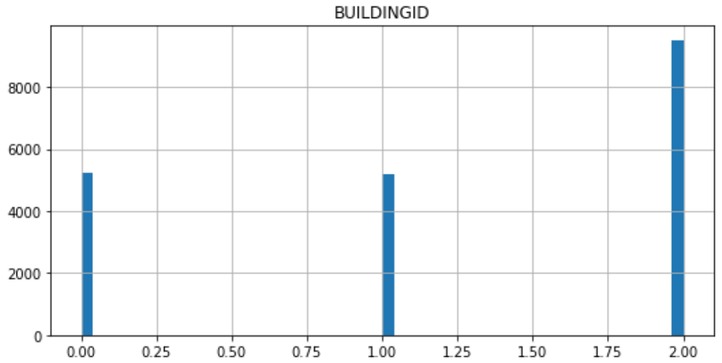
3. Remove "useless information"
4. Re- scale and Normalization
5. PCA

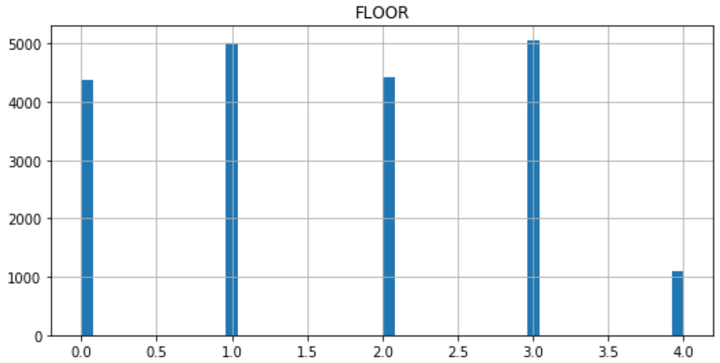
3 Modelling

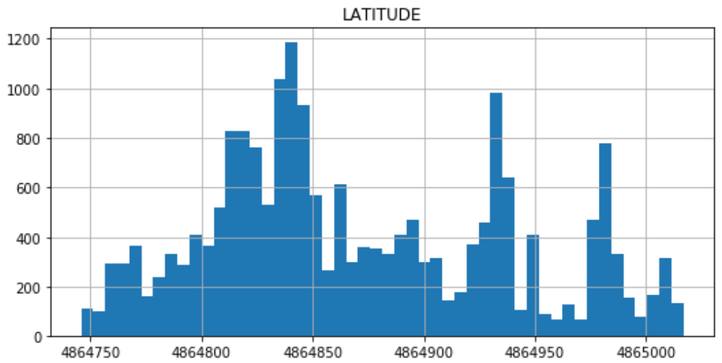
6. CV-Model performance — **Training Set**
7. Model Tuning — **Training Set**
8. Model Selection — **Validation Set**

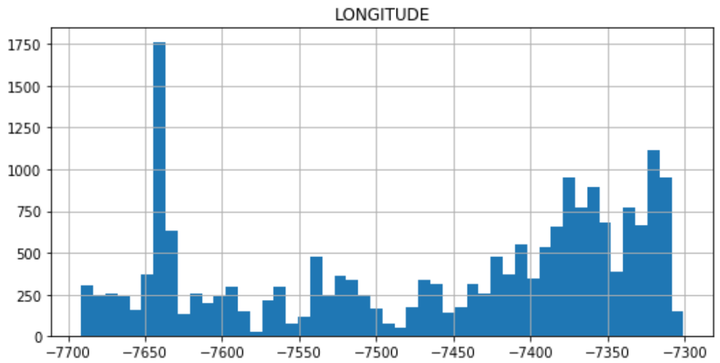
Data Exploration

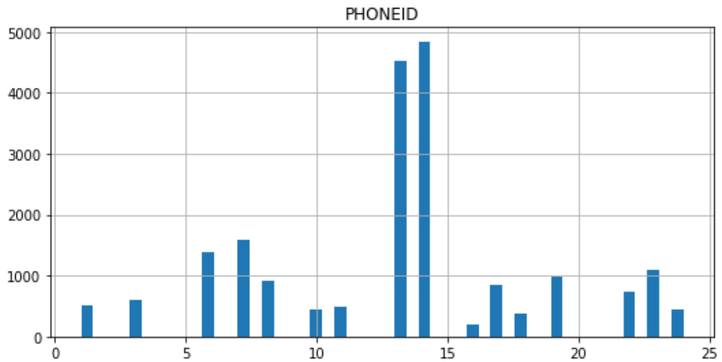
Distribution Graphs

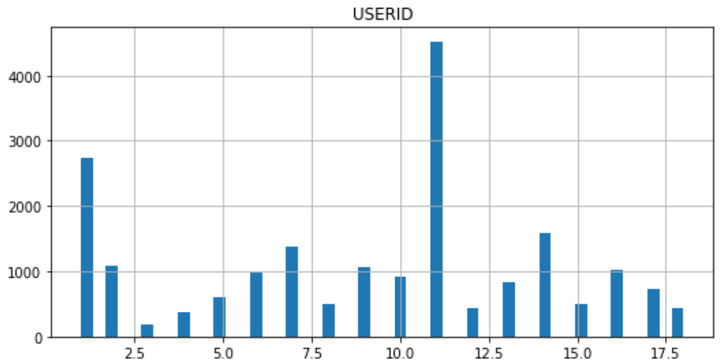






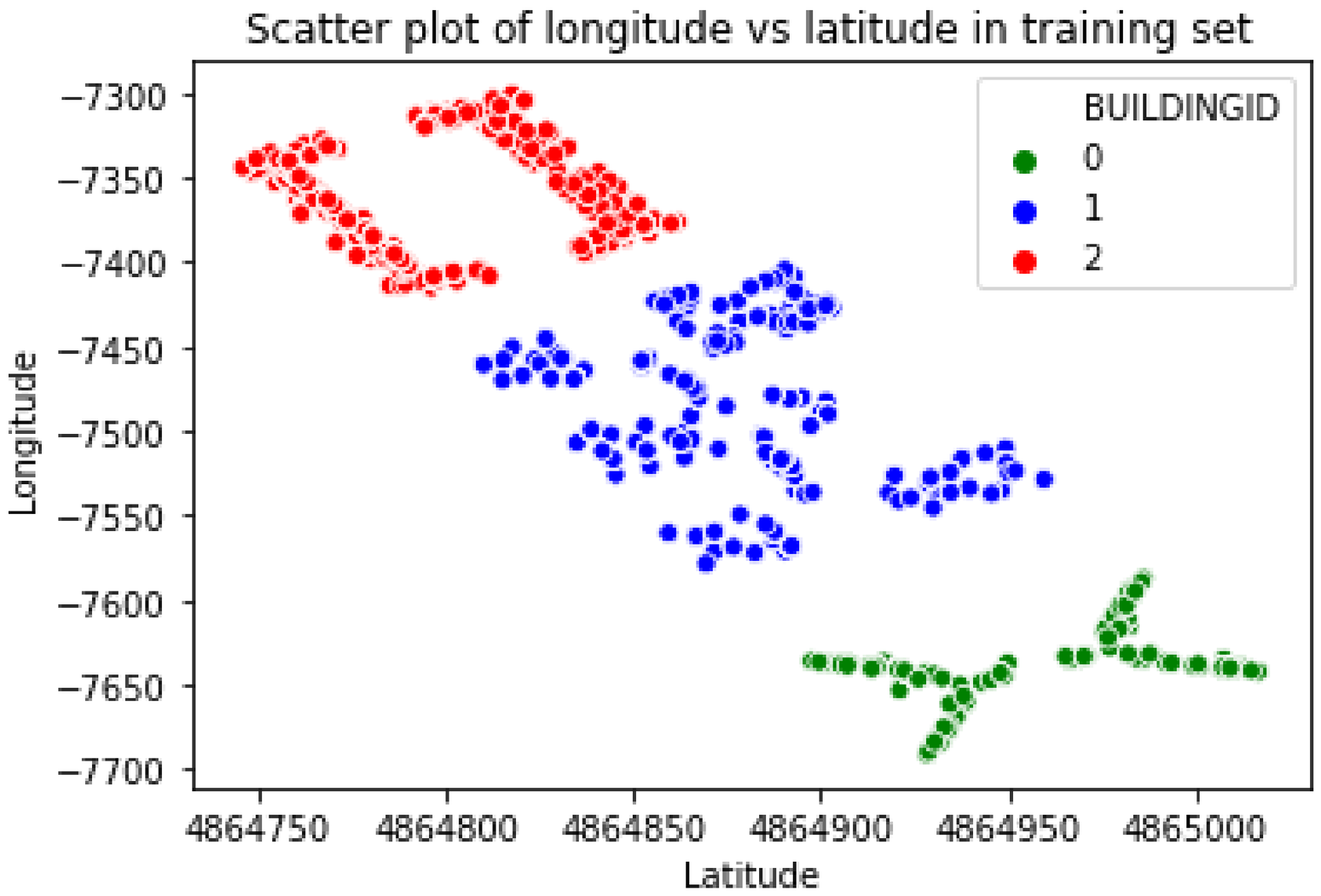






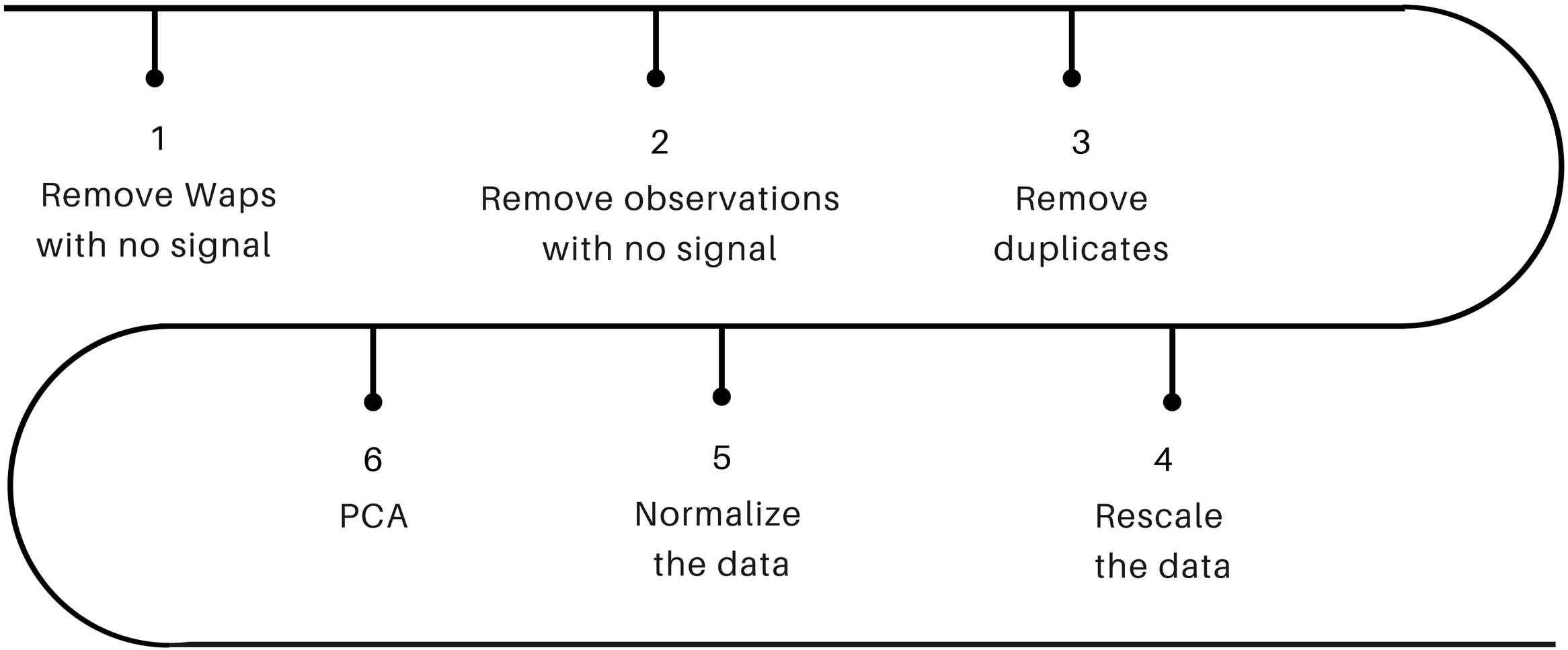
Data Exploration

Data Points Location



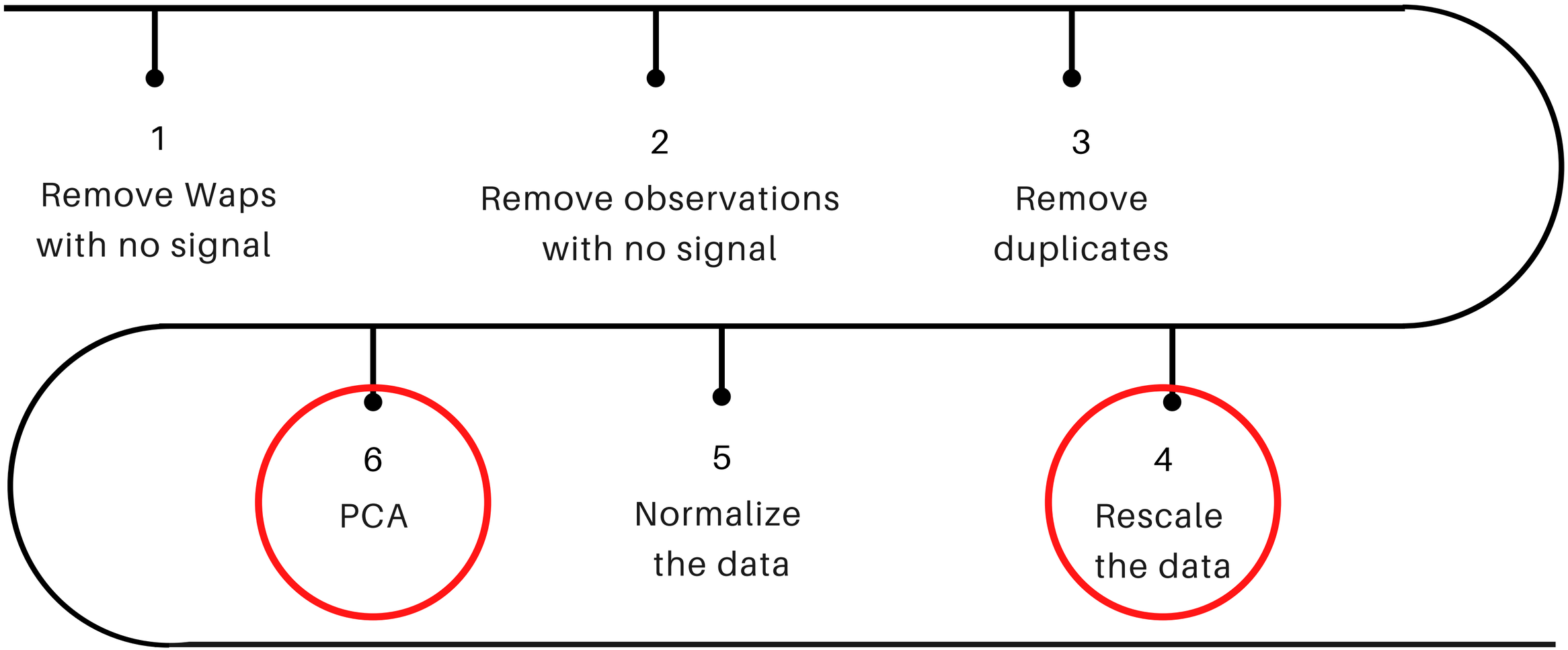
Data Pre-procesing

Preprocessing Flow



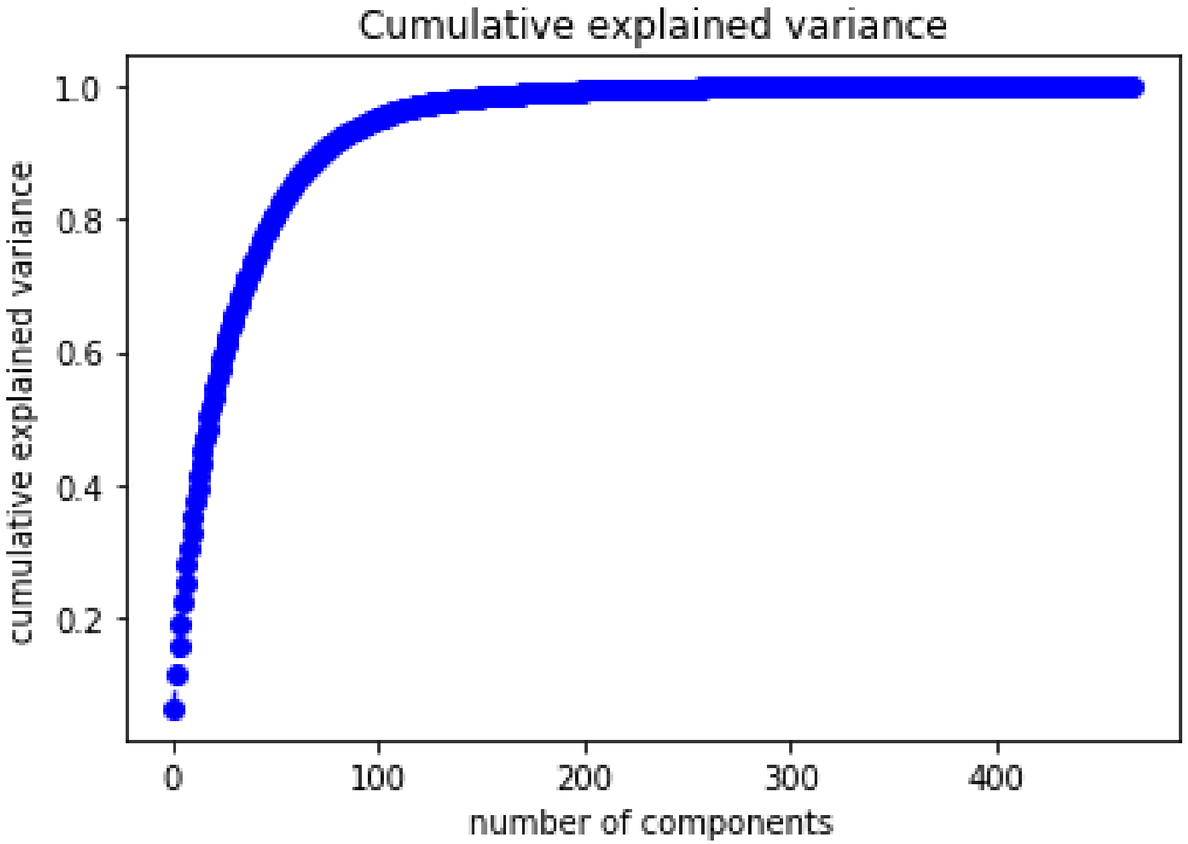
Data Pre-procesing

Preprocessing Flow



Data Pre-procesing

Differents Assumptions Considered



OPTION 1

- PCA= 100 components
- Re-scaling= Exponentiating
- Final Shape of WAPs= (19039, 100)

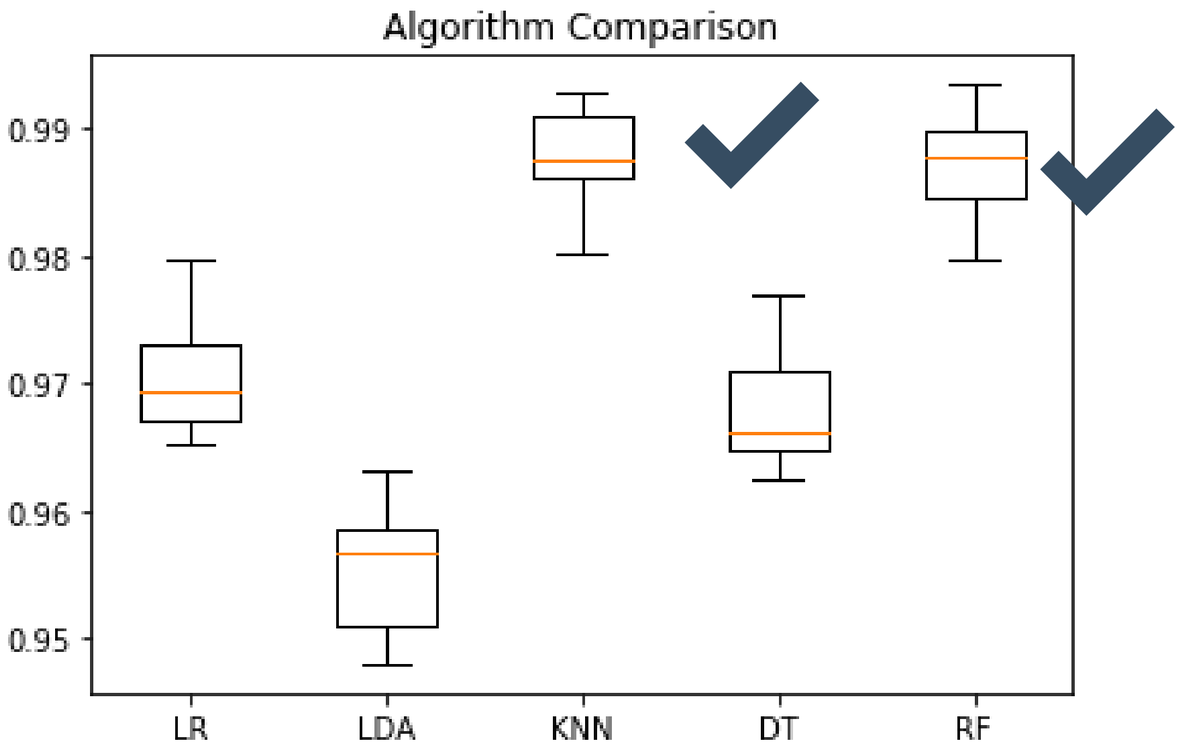
OPTION 2

- PCA= 99%
- Re-scaling= Positive values data representation.
- Final Shape of WAPs= (19039, 244)

Modelling

CV model performance

Floor



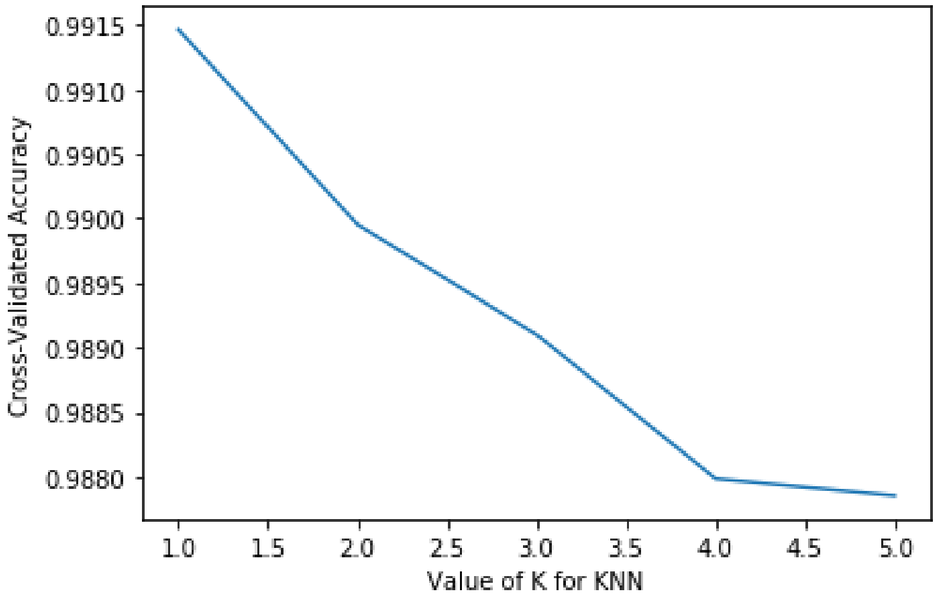
- **KNN and RF are the ones that perform better over the rest**

Training Set

Modelling

Model Tuning

Hyper parameter tuning-KNN



Hyper parameter tuning-RF

Parameters: {'n_estimators': 100, 'min_samples_split': 2, 'min_samples_leaf': 2, 'max_features': 'sqrt', 'max_depth': 100, 'bootstrap': False}

Training Set

Model Selection

KNN evaluation

Accuracy 0.949595 Kappa 0.929398

Confusion matrix

```
[[124   6     2    0    0]
[ 16   438    7    1    0]
[ 1     12   288   5    0]
[ 0      0    2   170   0]
[ 0      0    0    4   35]]
```

Random Forest evaluation

Accuracy 0.951395 Kappa 0.931686

Confusion matrix

```
[[119   11    2    0    0]
[ 10   441    9    2    0]
[ 1     11   292   2    0]
[ 0      0    2   169   1]
[ 0      0    0    3   36]]
```

Results

	BUILDING ID	FLOOR	LATITUDE	LONGITUDE
OPTION 1	Model: RF Accuracy: 100%	Model: RF Accuracy: 95%	Model: RF RMSE: 7.21	Model: RF RMSE: 7.94
OPTION 2	Model: LR Accuracy: 100%	Model: KNN Accuracy: 90%	Model: KNN RMSE: 9.48	Model: KNN RMSE: 9.29