BLACKWELL

Data Analytics

WIFI LOCATIONING

Agenda

Wifi-Locationing

1. Wifi Locationing – How does it work?

2. Wifi Locationing – Research Project University Jaume (Spain)

Data Analysis and ML for Wifi Locationing Data Set

1. Data Preparation

2. Data Exploration

3. Modeling & Predicting Location of User (**Building, Floor** & **Longitude/Latitude**)

Wifi-Locationing

GPS

- many applications need to know location of users
- outdoor localization problem solved accurately with

GPS sensors into the mobile devices

Wifi-Locationing

- indoor localization open problem due to loss of GPS signal
- estimating position of user (latitude, longitude and altitude)

by using WAPs and electronic device (mobile phone)

Wifi-Locationing by Univ. Jaume

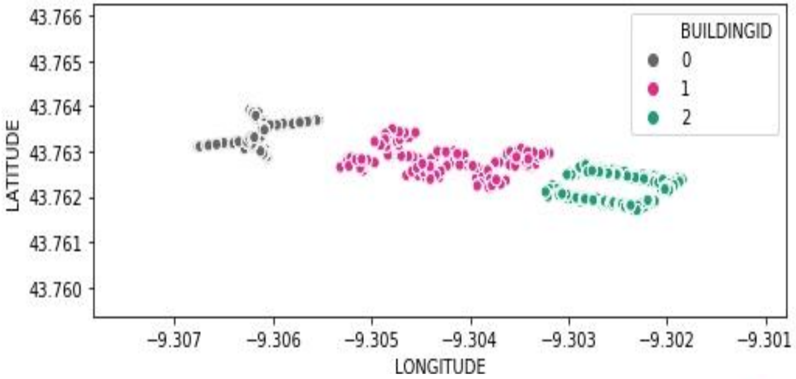
- 3 buildings on university campus

  (almost 110.000 m2)
- 3-4 floors per building

- 520 Wireless Access Points (WAPs)
- 20 different users with 25 Android devices

19.937 Wifi fingerprints for training

1.111 Wifi fingerprints for validation



Cleaning Wifi Data Set I

Raw Data Set

WAP010	...	WAP520	LONGITUDE	LATITUDE	FLOOR	BUILDINGID	SPACEID	RELATIVEPOSITION	USERID	PHONEID
100	...	100	-9.304907	43.763135	2	1	106	2	2	23
100	...	100	-9.304863	43.763258	2	1	106	2	2	23
100	...	100	-9.304661	43.763407	2	1	103	2	2	23
100	...	100	-9.304714	43.763265	2	1	102	2	2	23
100	...	100	-9.306088	43.763623	0	0	122	2	11	13

- 520 WAP's
- LONGITUDE
- LATITUDE
- FLOOR
- BUILDINGID
- SPACEID
- RELATIVEPOSTION
- USERID
- PHONEID
- TIMESTAMP

1. Step: Removing duplicate observations

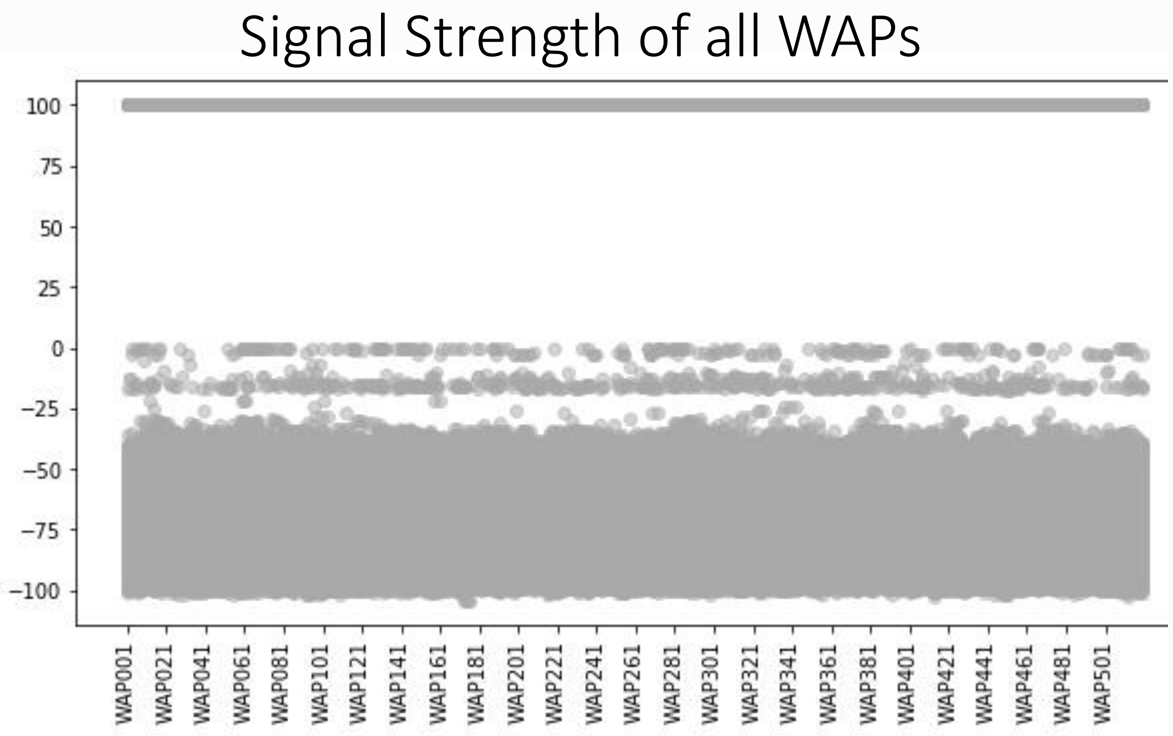
Observations	Features
19.937	529

Observations	Features
19.300	529

Cleaning Wifi Data Set II

2. Step: Converting Signal Strength for „no signal“

from +100 to -105 (lowest signal is-104)



3. Step: Removing redundant features

- WAPs with no signal

- irrelevant features

Observations	Features
19.300	529

→

Observations	Features
19.300	469

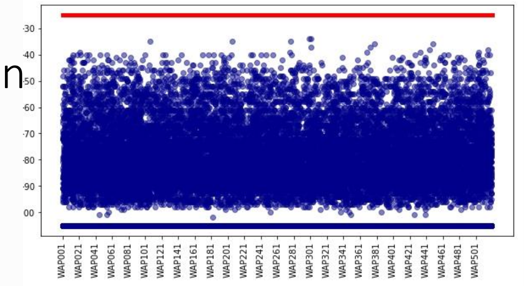
Cleaning Wifi Data Set III

4. Step: Removing Outliers

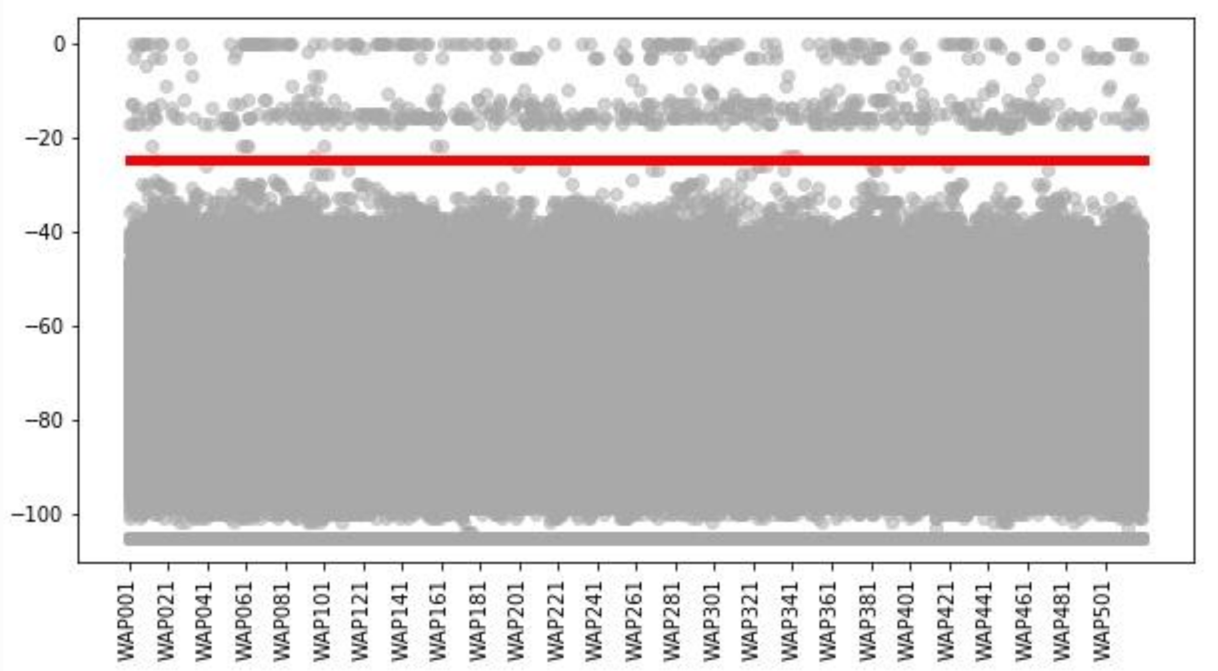
- definition of boundary by visualisation

- signal strength 0 until -25 outlier area

Validation Data Set



Signal Strength of all WAPs in training data set



-removing observations in this area

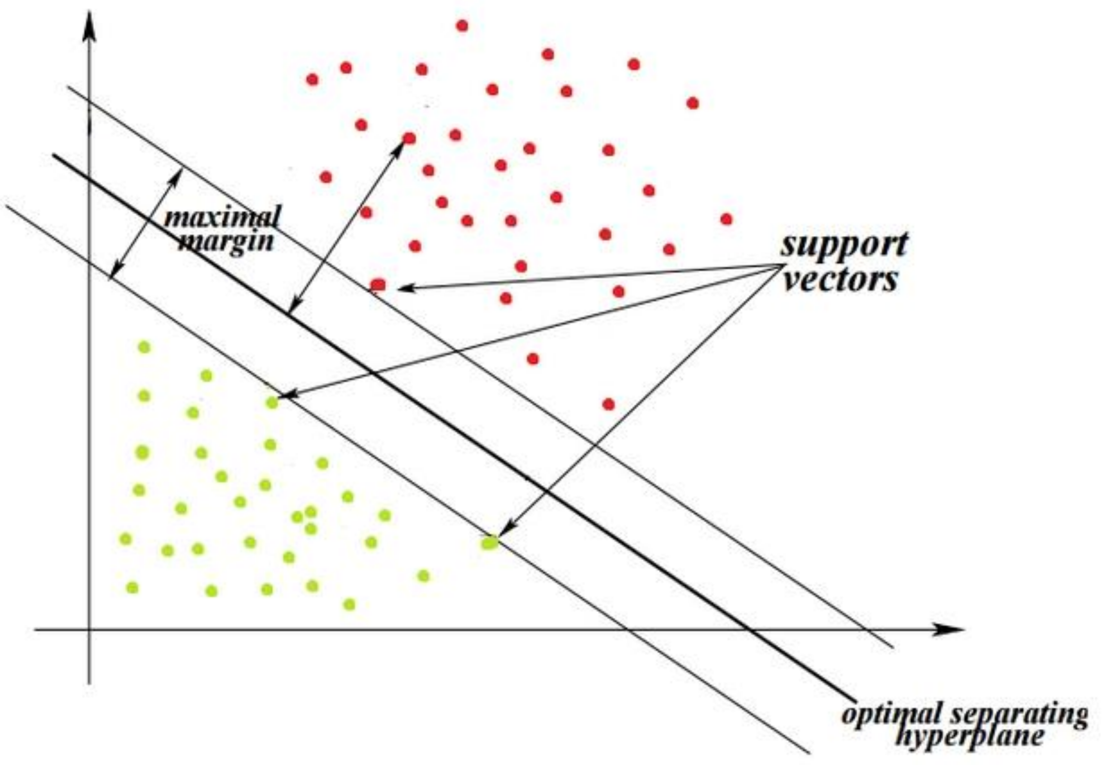
Observations	Features
19.300	469

Observations	Features
18.825	469

Predicting BUILDING ID I

Support Vector Machine

SVM: finds an optimal boundary between the possible outputs to identify the optimal separating hyperplane



grid search for best parameter combination for SVM:

C=5, gamma=0.001, kernel='rbf'

accuracy_score: 63.7 %

kappa_score: 46.9 %

Predicting BUILDING ID II

K Nearest Neighbor

KNN assumes similar things exist in close proximity

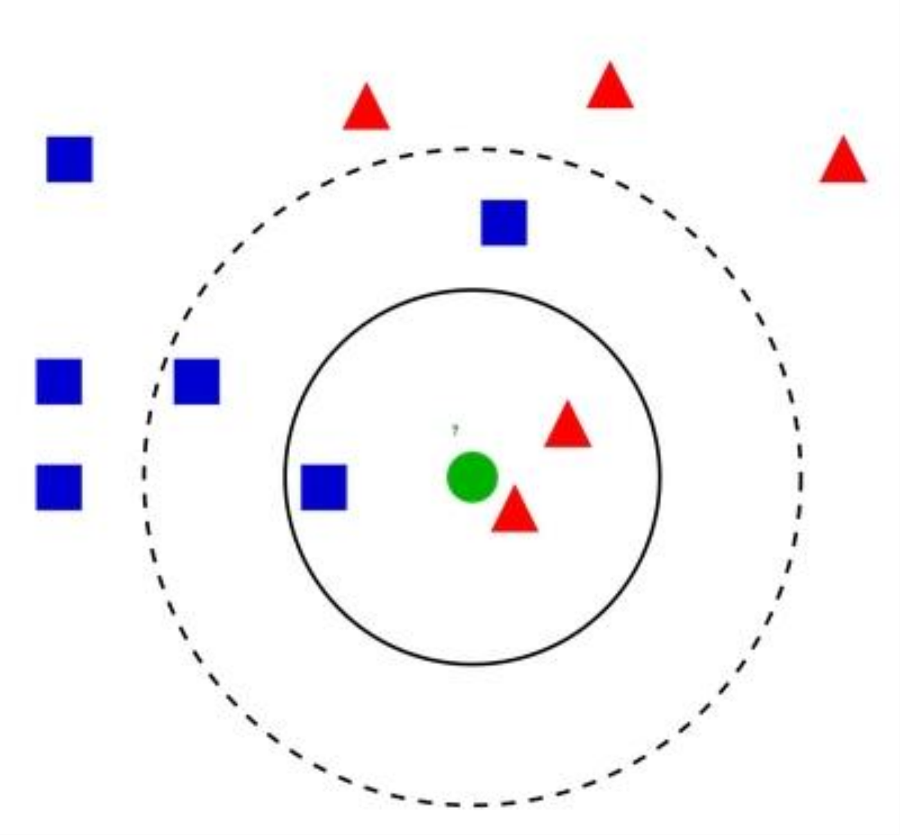
-calculating the distance between points

-majority of K's decide target class

grid search for best K : k_neighbors': 3

accuracy_score: 99.5 %

kappa_score: 99.3 %



Predicting BUILDING ID III

Random Forest

-constructing a multitude of decision trees

-what features allow to split observations so
  that the groups are as different as possible?

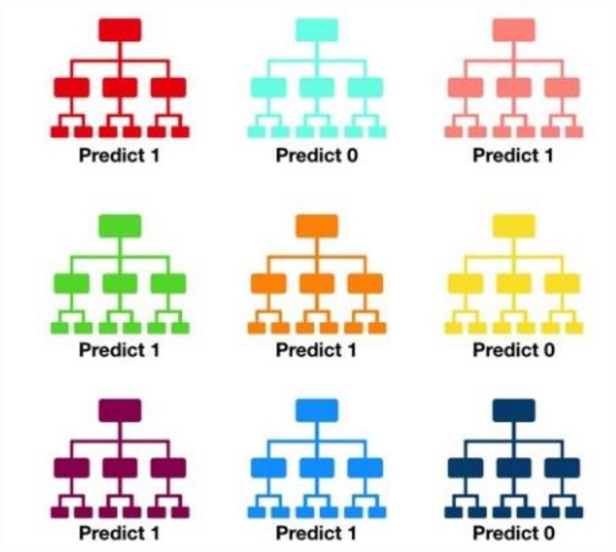
- Each tree gives prediction

- Class with most predictions -> model's prediction

accuracy_score: 99.8 %

kappa_score: 99.8 %

Random Forest



n_estimators': 50 for best model

Predicting 'FLOOR'

Random Forest

accuracy_score: 91.1 %

kappa_score: 87.6 %

<u>reasons for wrong predictions</u>:

-human error? confound floors?

-wrong placement of WAPs?

-> speculative

BUILDING 0

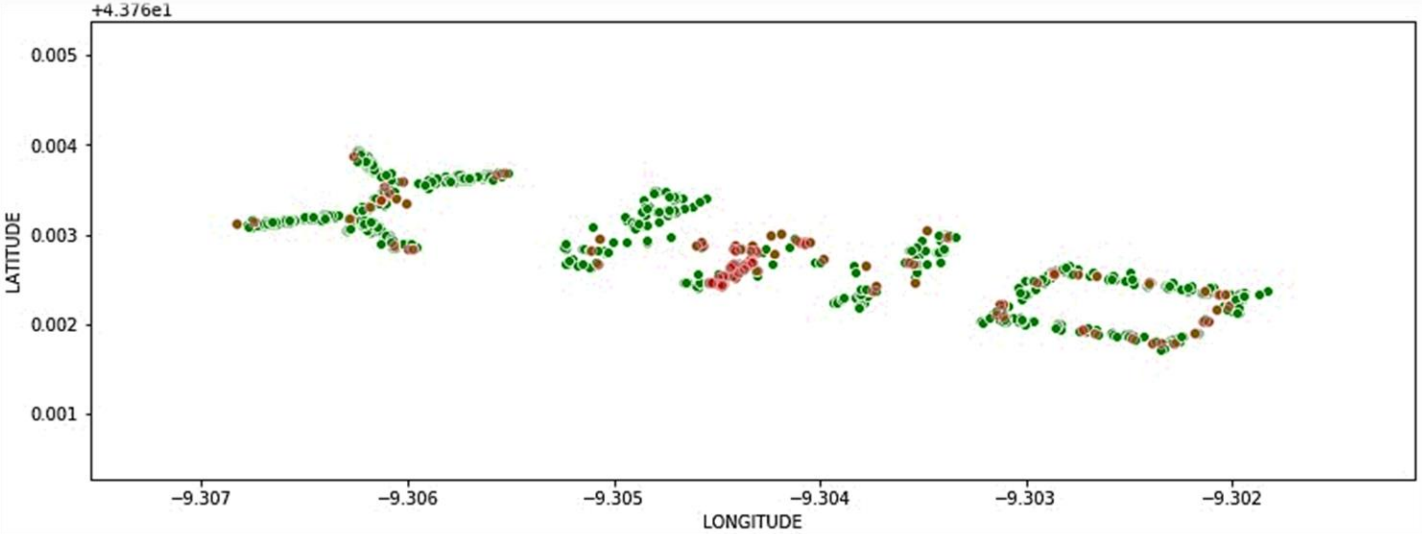
accuracy_score: 97.6 %

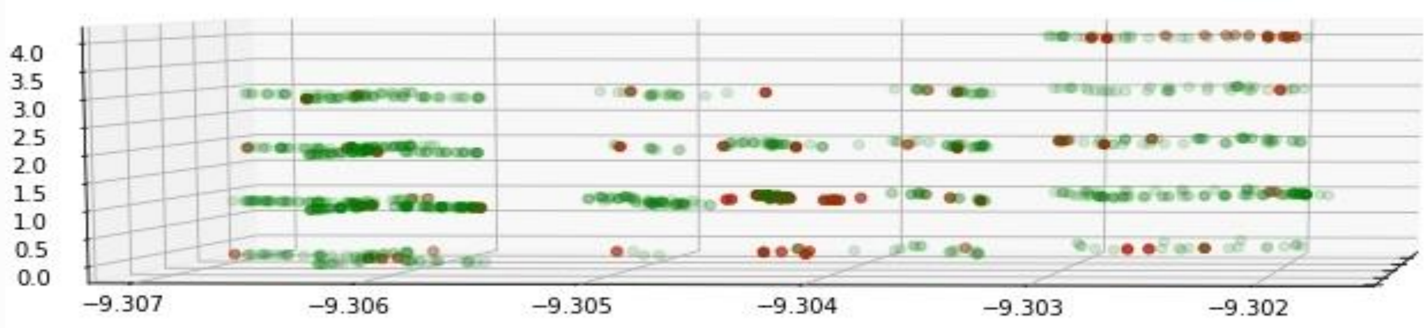
BUILDING 1

accuracy_score: 79.8 %

BUILDING 3

accuracy_score: 90.7 %





Predicting Longitude/Latitude

K-nearest neighbor (K=2)

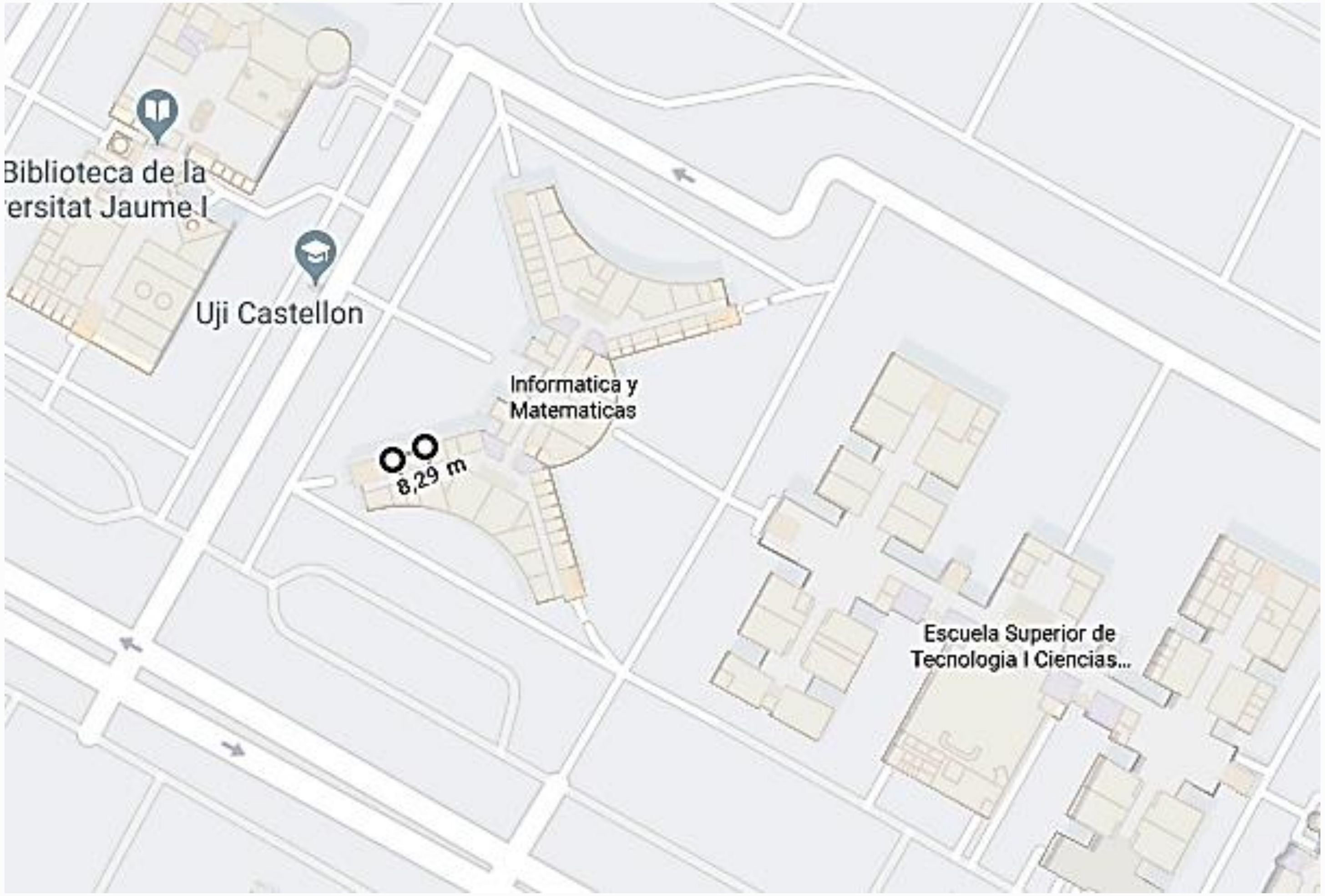
Longitude

mean absolute error: 6.20 m

Latitude

mean absolute error: 5.58 m

→ average location error: **8.34 m**



THANK **YOU**