Occupancy Prediction

Creative Data Mining

Natasha Catunda

HS 2017

Is it possible to accurately detect presence of people according to indoor environmental measurements in buildings?

Motivation

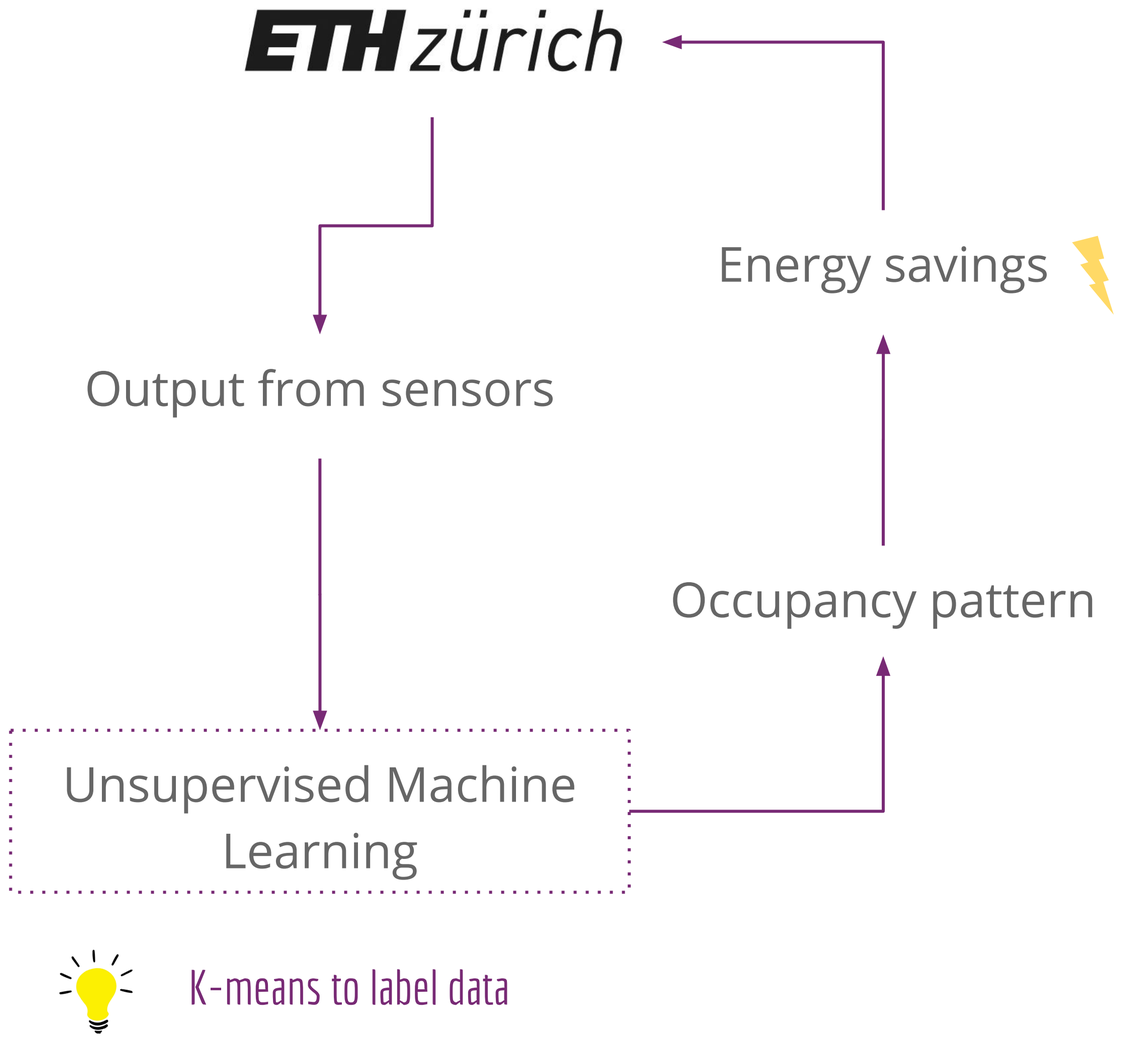
Accurate occupancy detection

HVAC control systems

Energy savings

*Candanedo, L. and Feldheim, V. (2016). **Accurate occupancy detection of an office room from light, temperature, humidity and CO 2 measurements using statistical learning models**. Energy and Buildings, 112, 15 January 2016, pp.28-39.

Motivation


ETH zürich
Energy savings
Output from sensors
Occupancy pattern
Unsupervised Machine Learning
K-means to label data

Dataset

1 room

9752 measurements

1 week

Every minute

Door open

Original dataset

	Index	Date	Temperature	Humidity	Light	CO2	HumidityRatio	Occupancy
0	1	2015-02-11 14:48:00	21.7600	31.133333	437.333333	1029.666667	0.005021	1
1	2	2015-02-11 14:49:00	21.7900	31.000000	437.333333	1000.000000	0.005009	1
2	3	2015-02-11 14:50:00	21.7675	31.122500	434.000000	1003.750000	0.005022	1
3	4	2015-02-11 14:51:00	21.7675	31.122500	439.000000	1009.500000	0.005022	1
4	5	2015-02-11 14:51:59	21.7900	31.133333	437.333333	1005.666667	0.005030	1

Inputs used

	Temperature	Humidity	Light	CO2	HumidityRatio
0	21.7600	31.133333	437.333333	1029.666667	0.005021
1	21.7900	31.000000	437.333333	1000.000000	0.005009
2	21.7675	31.122500	434.000000	1003.750000	0.005022
3	21.7675	31.122500	439.000000	1009.500000	0.005022
4	21.7900	31.133333	437.333333	1005.666667	0.005030

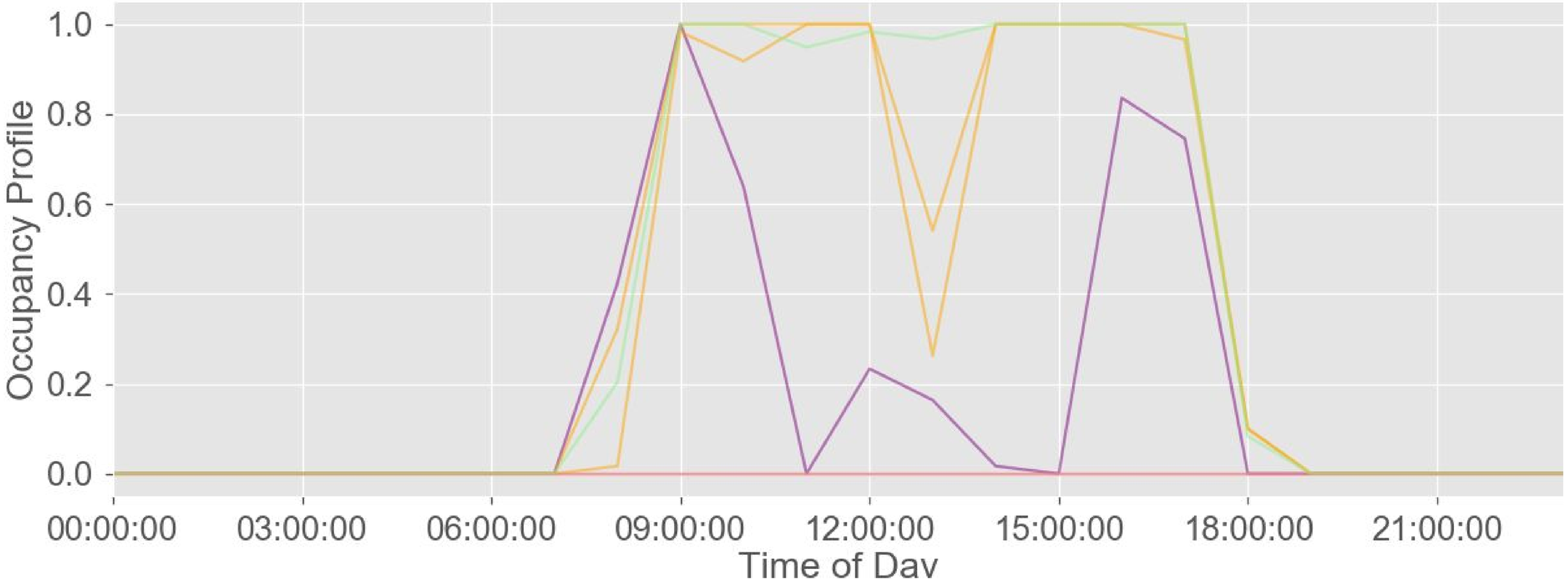
Normalized inputs

	Temperature	Humidity	Light	CO2
0	0.462168	0.525565	0.276618	0.342373
1	0.468303	0.518004	0.276618	0.323736
2	0.463701	0.524950	0.274510	0.326092
3	0.463701	0.524950	0.277672	0.329704
4	0.468303	0.525565	0.276618	0.327296

Data set from Candanedo, L. and Feldheim, V. (2016).

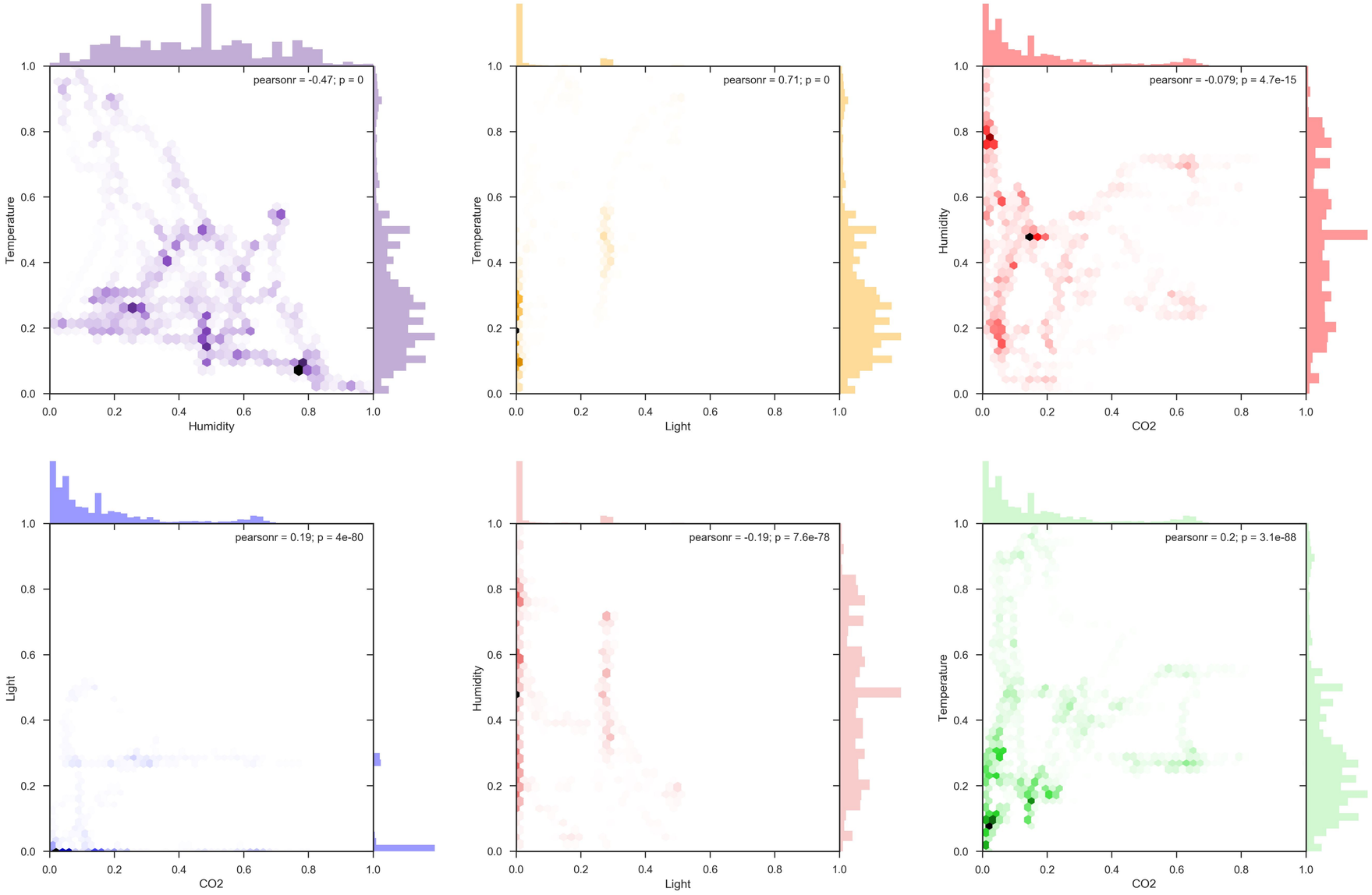
Dataset

Daily profiles
original data



Data set from Candanedo, L. and Feldheim, V. (2016).

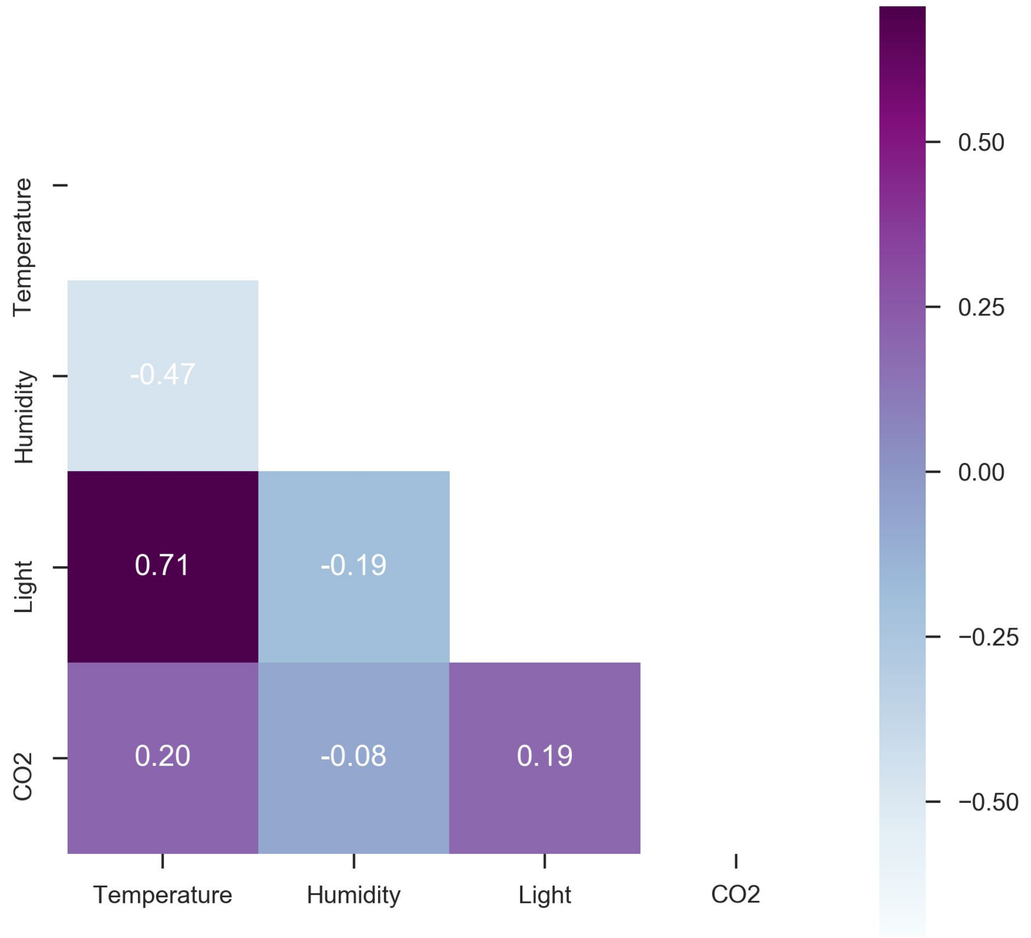
Hexagonal binning plots



Data set from Candanedo, L. and Feldheim, V. (2016).

Data analysis

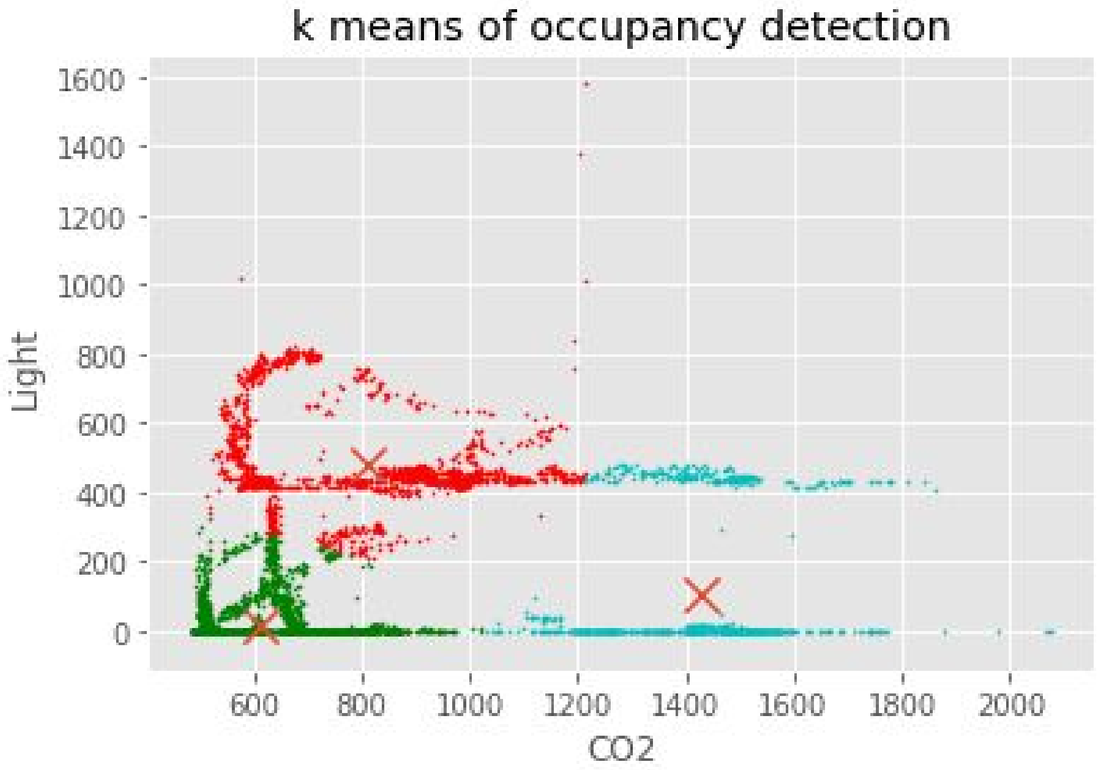
Correlation matrix



Data analysis

Data set from Candanedo, L. and Feldheim, V. (2016).

K-means clustering



- 3 clusters
- Silhouette score of 0.67

Results

Comments

	Temperature	Humidity	Light	CO2	HumidityRatio
0	21.7600	31.133333	437.333333	1029.666667	0.005021
1	21.7900	31.000000	437.333333	1000.000000	0.005009
2	21.7675	31.122500	434.000000	1003.750000	0.005022
3	21.7675	31.122500	439.000000	1009.500000	0.005022
4	21.7900	31.133333	437.333333	1005.666667	0.005030

Silhouette score of 0.32 !

Door open

Door closed

Silhouette score of 0.67

Silhouette score of 0.77

Thanks! Questions?

Creative Data Mining

Natasha Catunda

HS 2017

Appendix

Data set	Nº of observations	Variables	Obs.
Training	8143	7	Door closed
Test 1	2665	7	Door closed
Test 2	9752	7	Door open

The public rooms used by students in Hönggerberg usually have their door open.