



MSC THESIS DEFENSE

Quang Phong – 6286943

Detect and estimate occupancy via IoT-enabled air quality sensor with machine learning – a potential privacy-preserving approach to deal with looming Covid-19 waves and winter gas shortage



01



ABOUT ME
AND MY COMPANY

02

RATIONALE
AND RESEARCH QUESTION

03

THEORIES, EMPIRICAL
STUDIES AND GAPS

04

WHAT I'VE DONE

05

WHAT I'VE FOUND

06

DISCUSSION

07

RECOMMENDATIONS





01



ABOUT ME
AND MY COMPANY

02

RATIONALE
AND RESEARCH QUESTION

03

THEORIES, EMPIRICAL
STUDIES AND GAPS

04

WHAT I'VE DONE

05

WHAT I'VE FOUND

06

DISCUSSION

07

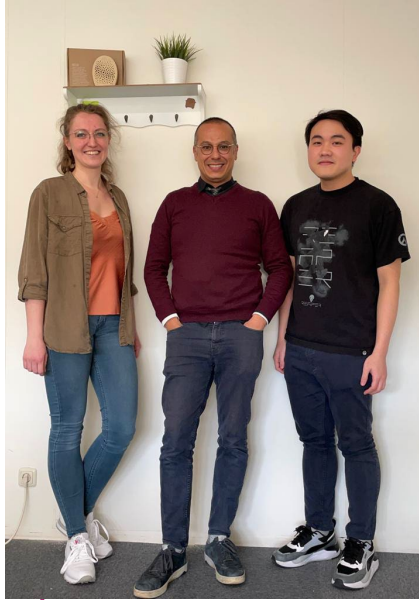
RECOMMENDATIONS





ABOUT ME

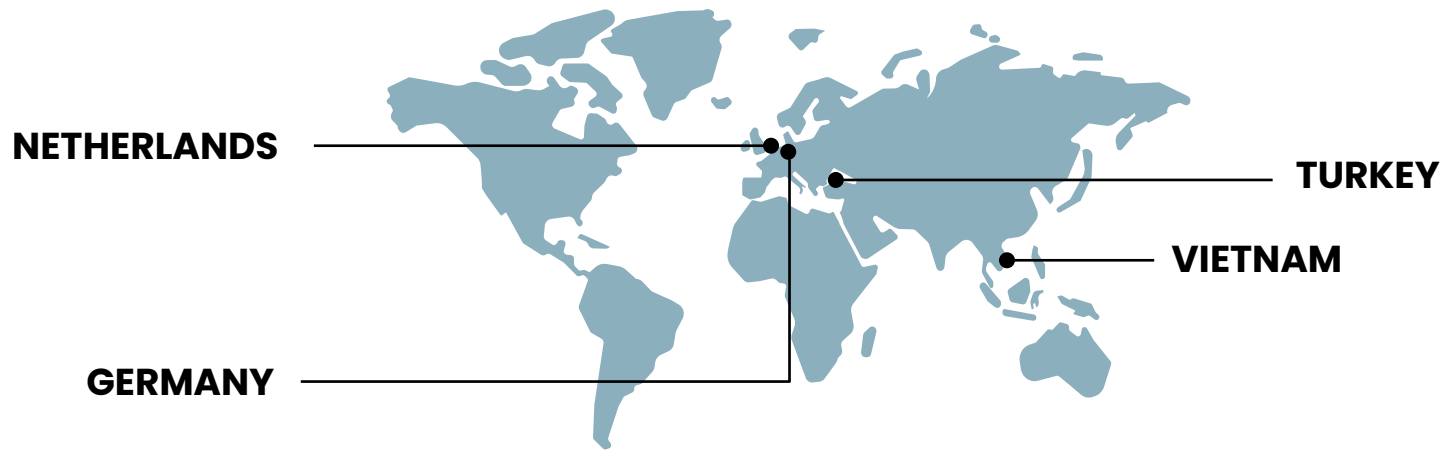
- Quang Trieu Nhat Phong
("Millions of sunbeams")
- MSc. Business Intelligence
and Smart Services
- Ho Chi Minh City, Vietnam



sustainably.io

- start-up aimed at sustainability
- offer air quality measurement
- use biodegradable materials
- cooperate with partners that employ vulnerable people

sustainably.io





A quiet place (2018)

Don't make noise!

Don't look!

Bird box (2018)





01



ABOUT ME
AND MY COMPANY

02

RATIONALE
AND RESEARCH QUESTION

03

THEORIES, EMPIRICAL
STUDIES AND GAPS

04

WHAT I'VE DONE

05

WHAT I'VE FOUND

06

DISCUSSION

07

RECOMMENDATIONS



RATIONALE

Energy crisis

The crisis facing us in winter (and already in summer with the rising price of energy).

Covid-19

We are not yet to live in an post-Coronavirus era.

GDPR

This gives everyone a right say “no” to privacy-invasive sensors.





RESEARCH QUESTION

“Is it possible to detect and estimate occupancy from air sensors in SBE?”

RESEARCH OBJECTIVES

01

What level of occupancy can we derive from air data: binary detection, multi-class classification, or numerical estimation?

02

What algorithms and variables contribute to good occupancy detection modelling?

03

What are the requirements to successfully detect occupancy in an indoor place like SBE staff lounge?





01



ABOUT ME
AND MY COMPANY

02

RATIONALE
AND RESEARCH QUESTION

03

THEORIES, EMPIRICAL
STUDIES AND GAPS

04

WHAT I'VE DONE

05

WHAT I'VE FOUND

06

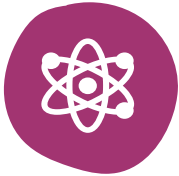
DISCUSSION

07

RECOMMENDATIONS



AIR SENSOR DATA



PM



CO2



TEMPERATURE



**RELATIVE
HUMIDITY**

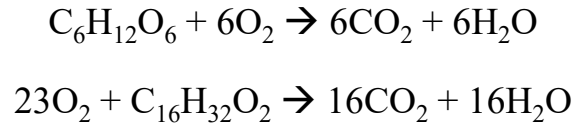


AIR PRESSURE

RELATIONSHIP OF HUMAN AND AIR COMPONENTS



CO₂



TEMPERATURE

MASS BALANCE OF CO₂ EQUATION



CO₂

$$V * \frac{dC}{dt} = E + Q(C_o - C_i)$$



KEY RESEARCH GAPS

Privacy-preserving

not camera, sound,
and heat sensors

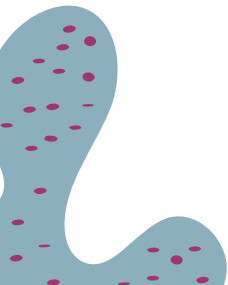
Large room, multi-class

Multiple air variables

Ensemble model

XGBoost, Random
Forest

HVAC-equipped room





01



ABOUT ME
AND MY COMPANY

02

RATIONALE
AND RESEARCH QUESTION

03

THEORIES, EMPIRICAL
STUDIES AND GAPS

04

WHAT I'VE DONE

05

WHAT I'VE FOUND

06

DISCUSSION

07

RECOMMENDATIONS

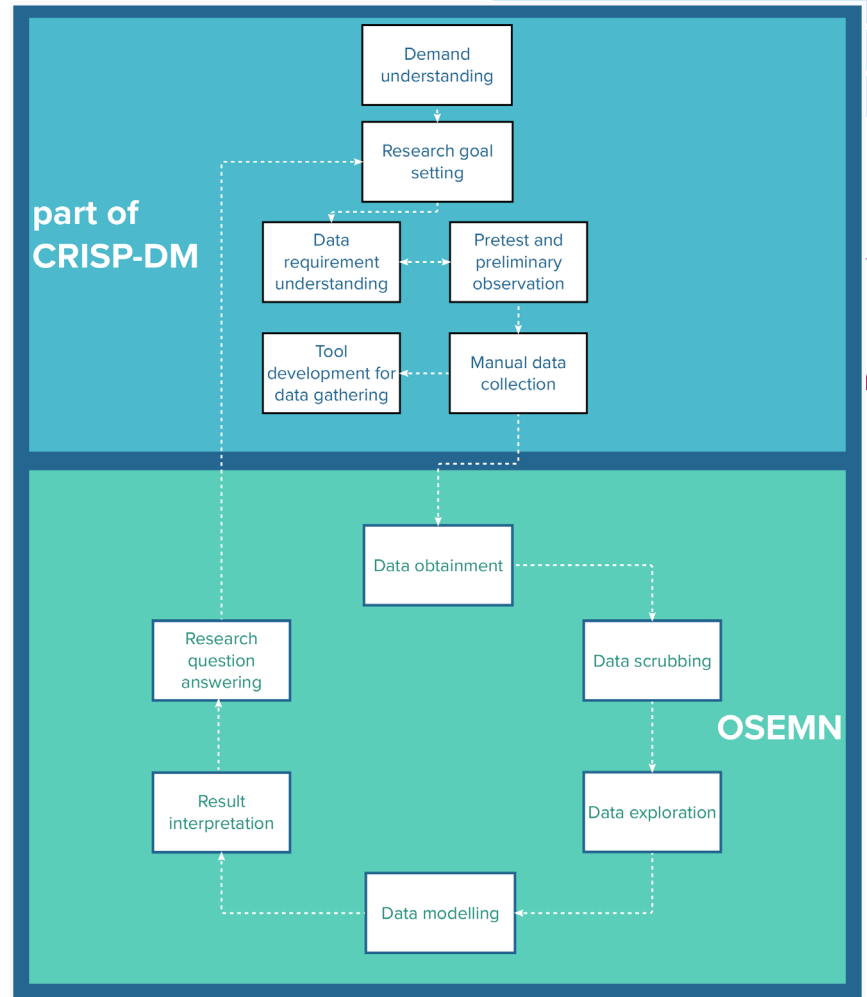


RESEARCH DESIGN

June 16 to July 11, 2022.

4552 minutely observations from 12 days.

On weekdays and during working hours.





TOOL DEVELOPMENT

How it works?

People counter with 2 active IR sensors to detect people passing the device

Why it needs enhancement?

They are sensitive to sunlight, to arm swings, unable to detect groups



MACHINE LEARNING

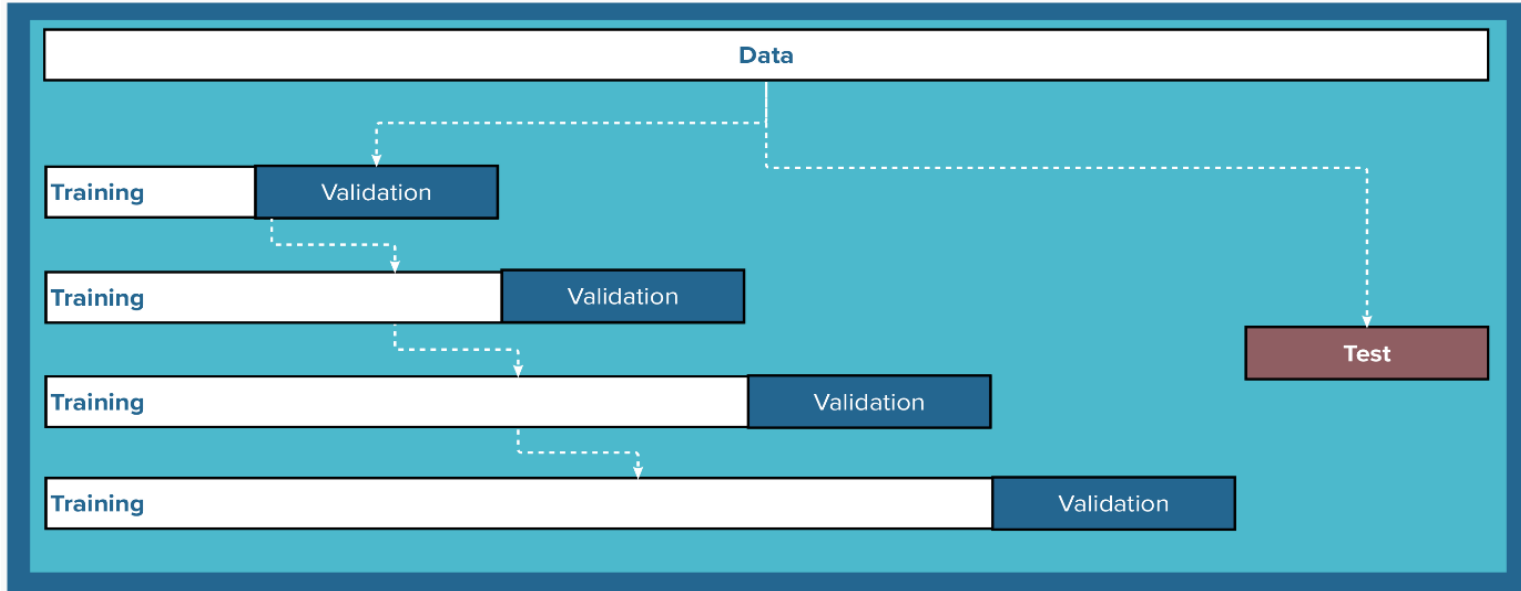
Models

Logistic regression
Support Vector Machine
Random Forest
Gradient Extreme Boosting

ML concepts

Train-validation-test
5-fold rolling validation
Grid Search hyperparameter
optimization

MACHINE LEARNING





01



ABOUT ME
AND MY COMPANY

02

RATIONALE
AND RESEARCH QUESTION

03

THEORIES, EMPIRICAL
STUDIES AND GAPS

04

WHAT I'VE DONE

05

WHAT I'VE FOUND

06

DISCUSSION

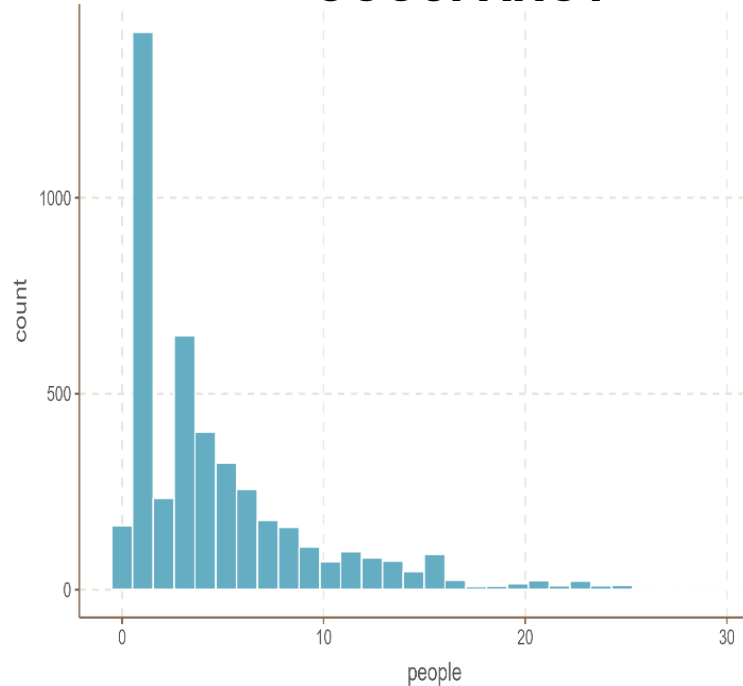
07

RECOMMENDATIONS

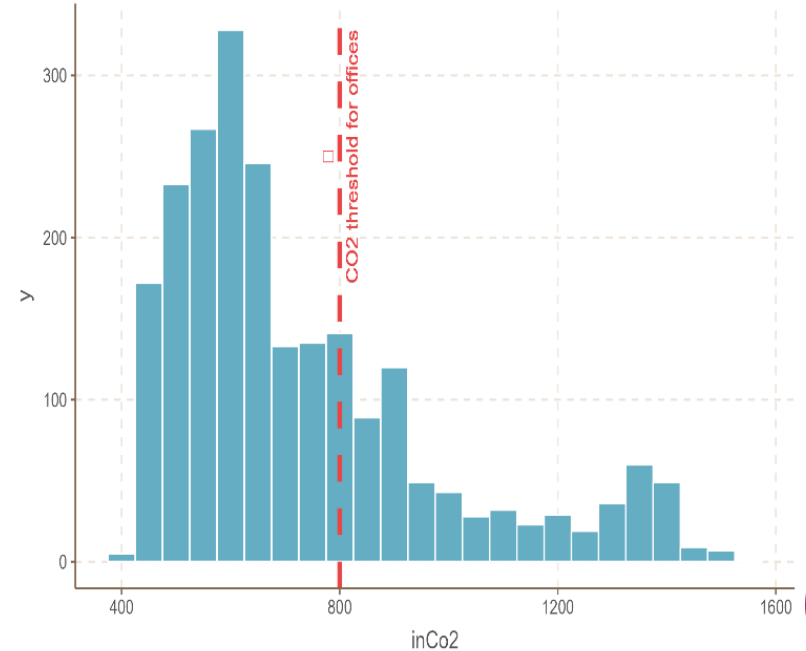


DESCRIPTIVE ANALYSIS

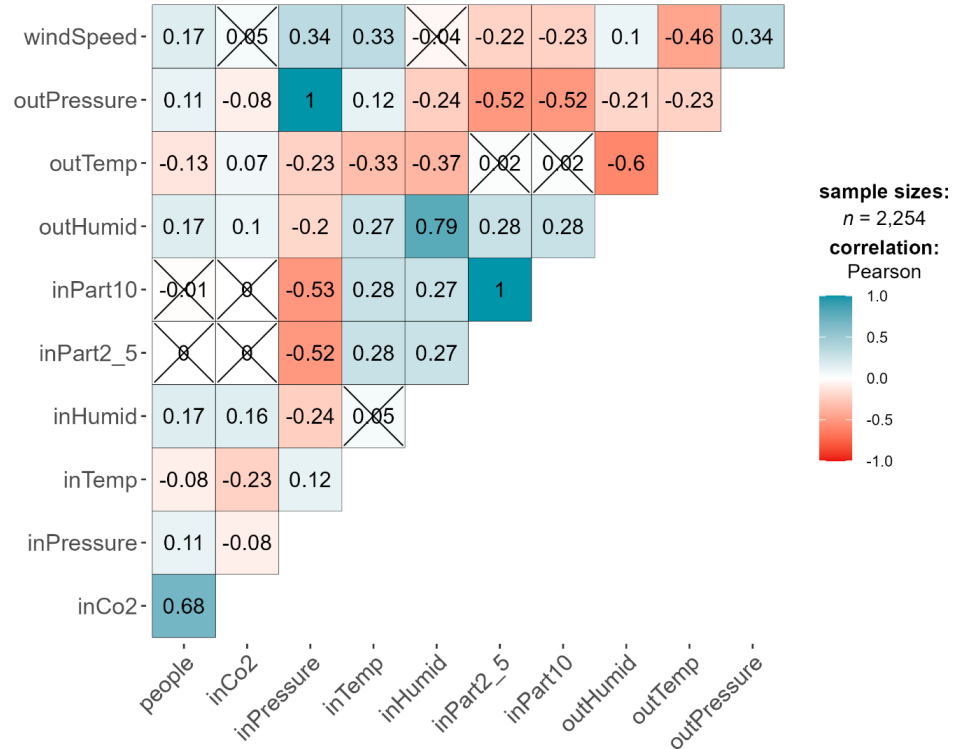
OCCUPANCY



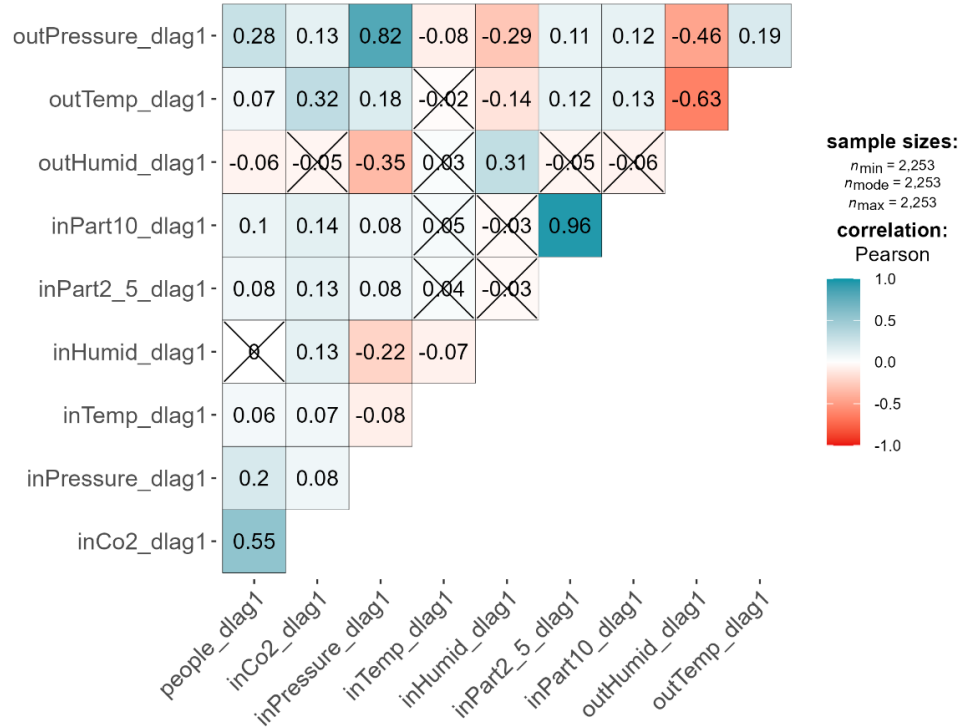
CO2



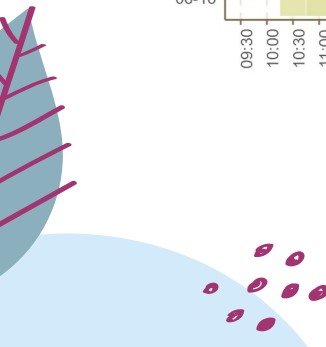
CORRELATION



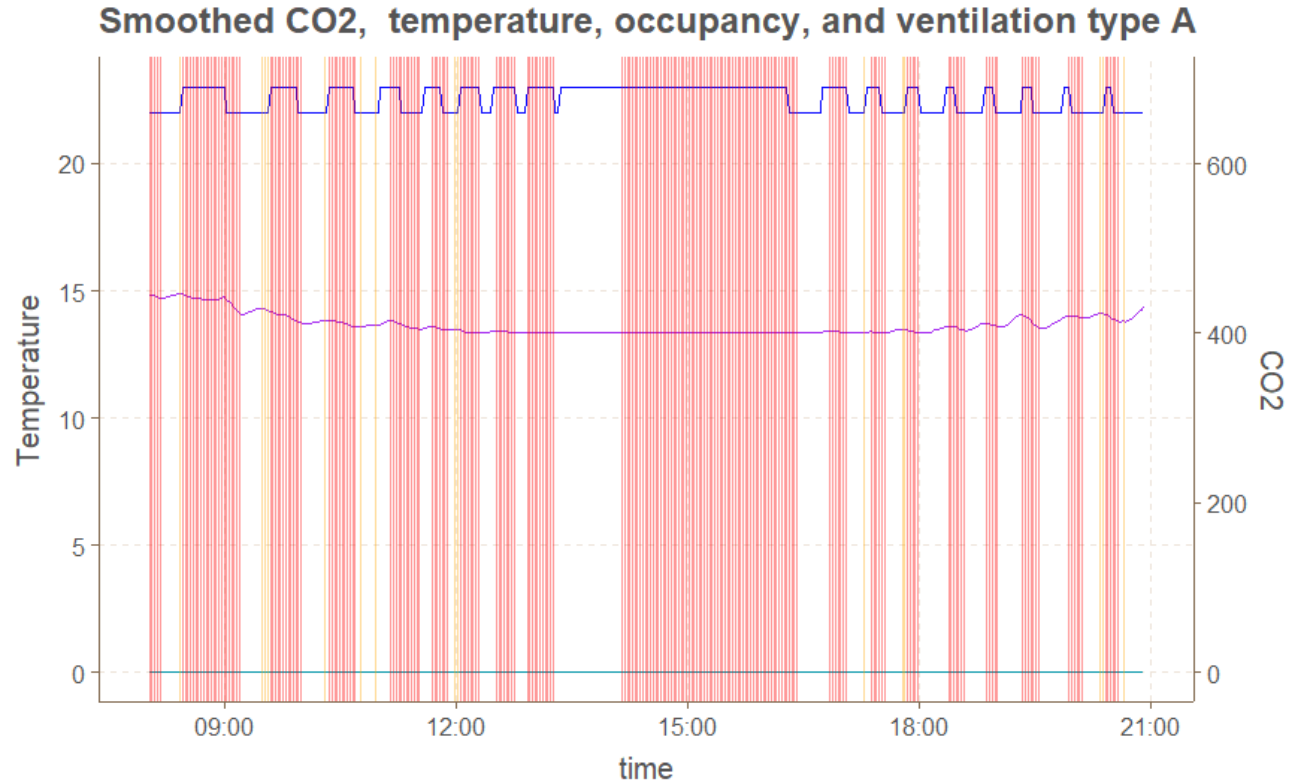
CORRELATION



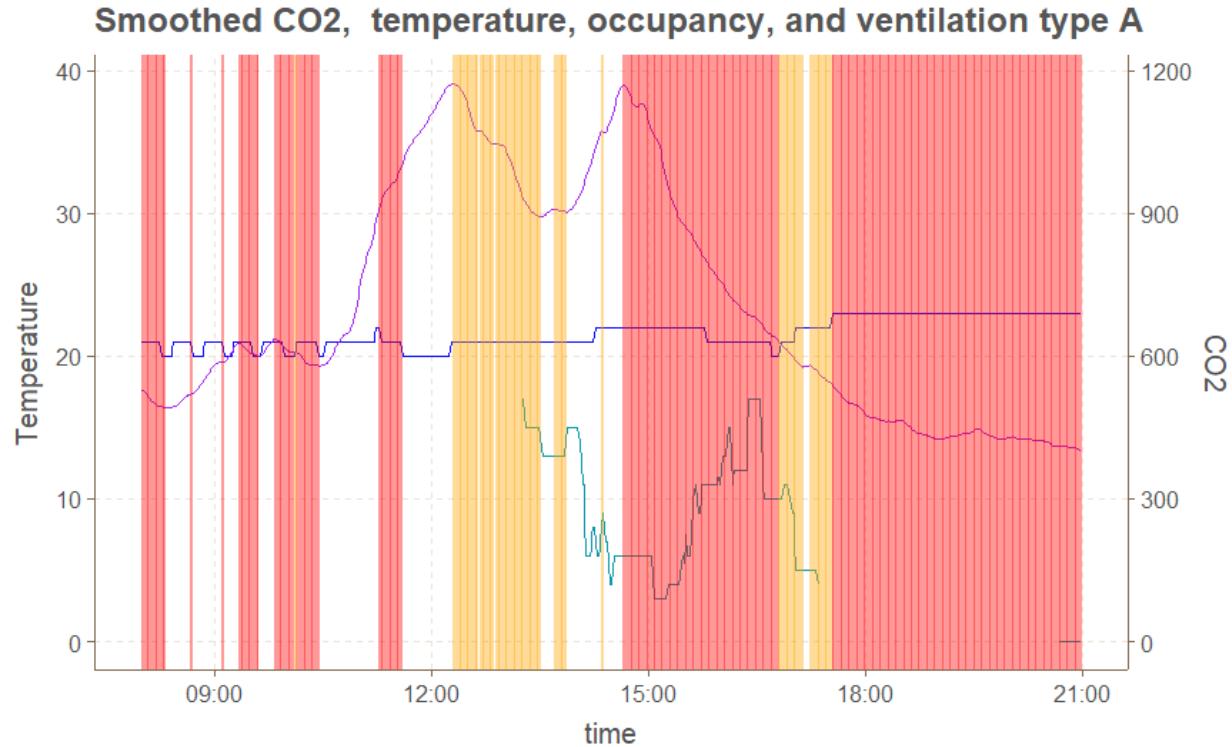
X = non-significant at $p < 0.05$ (Adjustment: Holm)



DETECT VENTILATION

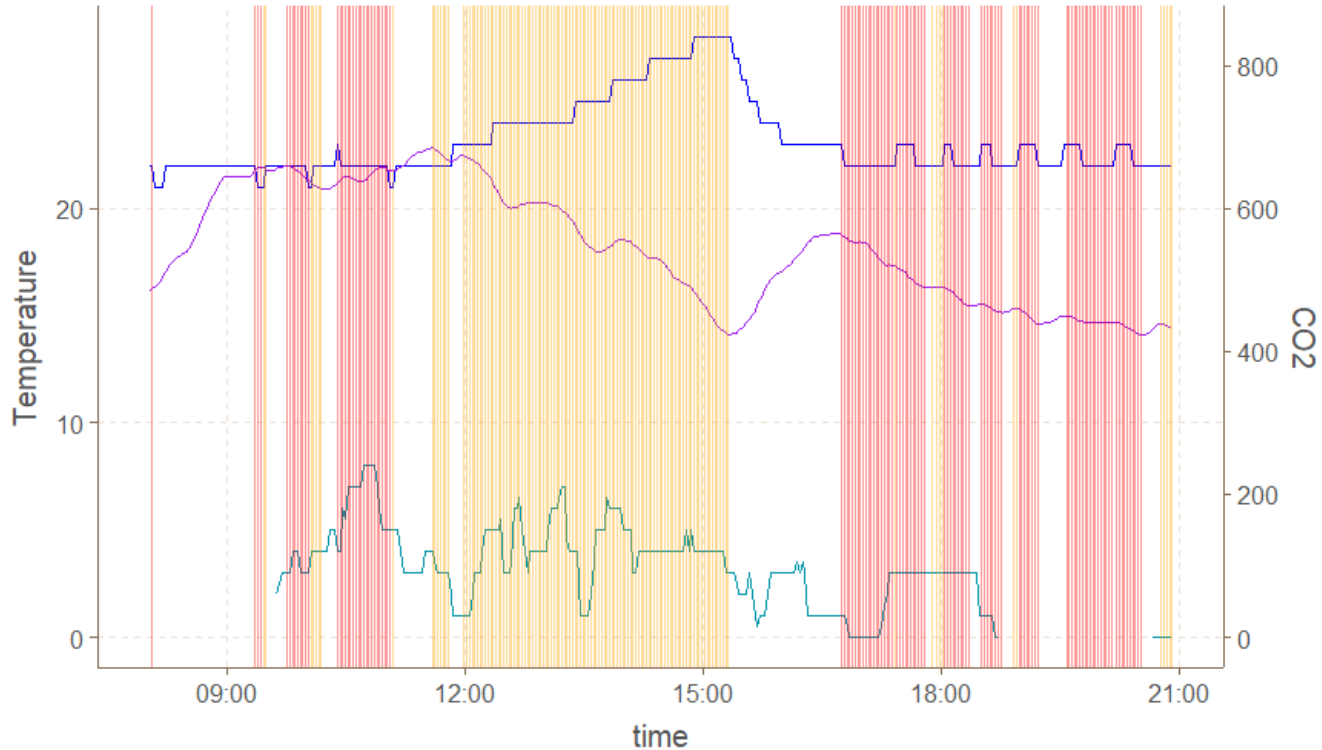


DETECT VENTILATION

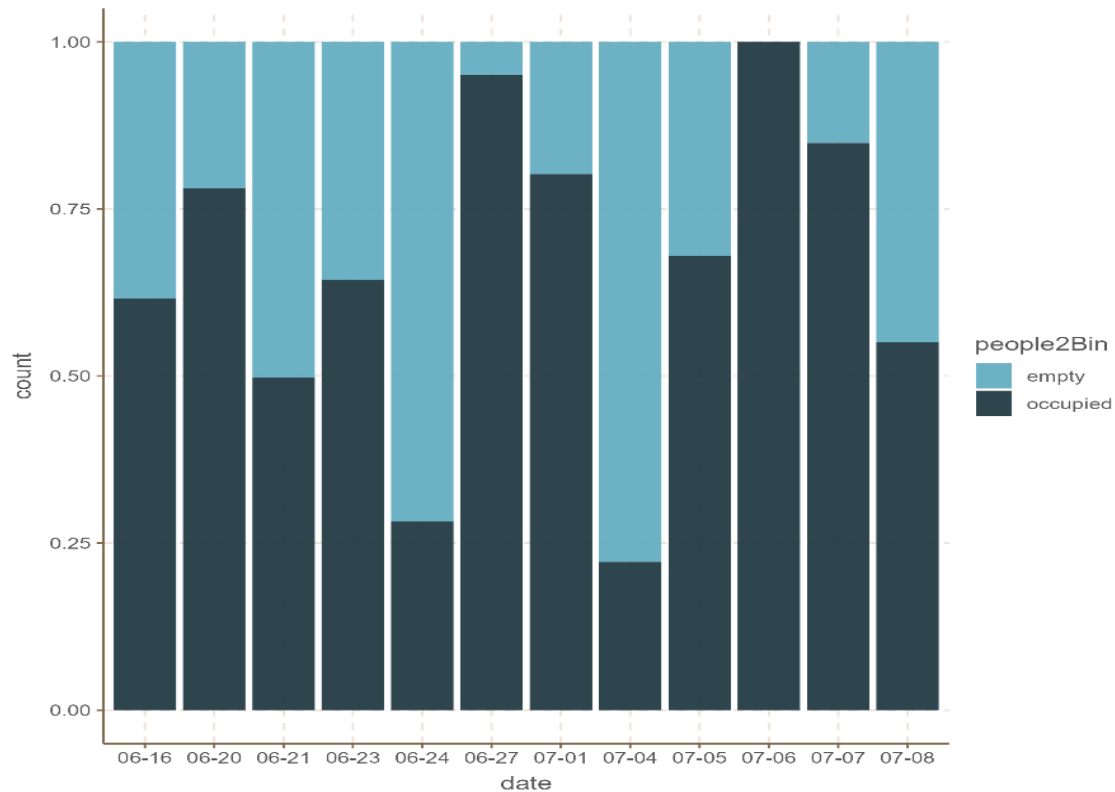


DETECT VENTILATION

Smoothed CO2, temperature, occupancy, and ventilation type A



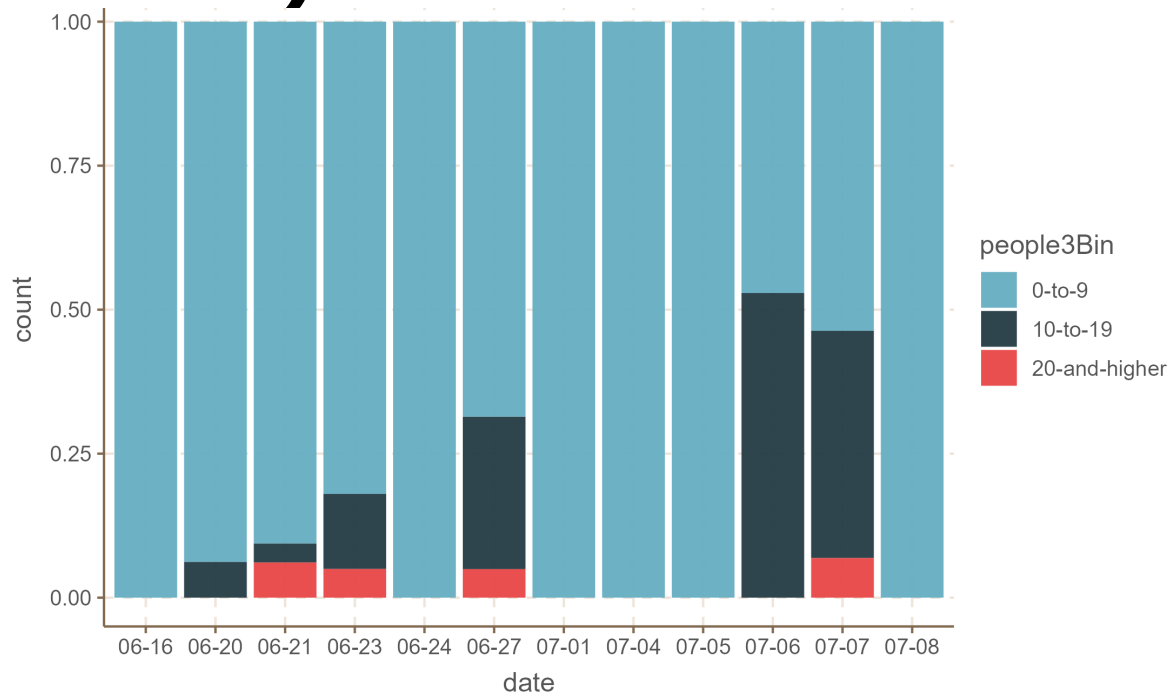
OCCUPANCY DETECTION



OCCUPANCY DETECTION

Indoor CO2, ventilation	Accuracy	Sensitivity	Specificity	Precision
XGBoost	83.24%	84.19%	80.77%	91.97%
RF	77.39%	76.10%	80.77%	91.19%
Logistic regression	81.91%	78.68%	90.38%	95.54%
SVM	81.65%	85.29%	71.55%	88.55%

MULTI-CLASS CLASSIFICATION (imbalanced)



MULTI-CLASS CLASSIFICATION (imbalanced)

Without periodical validation data

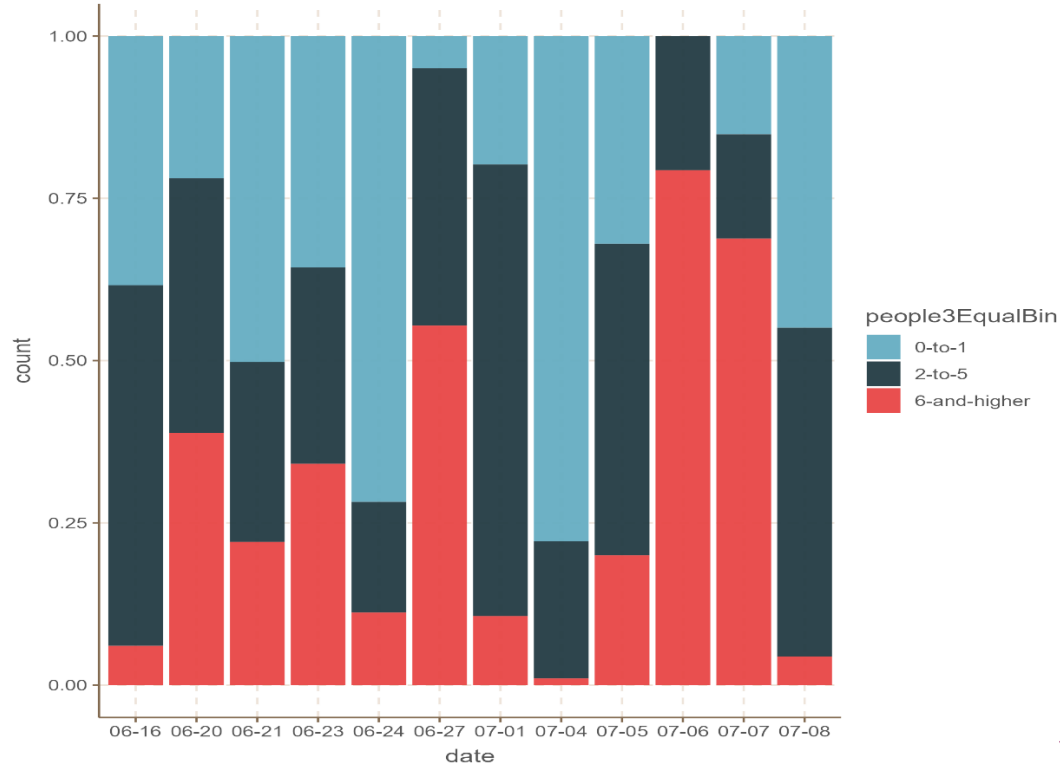
Accuracy (70%)	Not crowded	Moderate	Crowded
Not crowded	248	49	0
Moderate	0	0	0
Crowded	27	37	15

MULTI-CLASS CLASSIFICATION (imbalanced)

With periodical validation data

Accuracy (93.62%)	Not crowded	Moderate	Crowded
Not crowded	261	8	0
Moderate	14	78	2
Crowded	0	0	13

MULTI-CLASS CLASSIFICATION (balanced)



MULTI-CLASS CLASSIFICATION (balanced)

Without periodical validation data

Accuracy (67.8%)	0-1	2-5	6-or-more
0-1	92	12	0
2-5	52	45	18
6-or-more	3	46	118

MULTI-CLASS CLASSIFICATION (balanced)

With periodical validation data

Accuracy (72.9%)	0-1	2-5	6-or-more
0-1	73	31	0
2-5	25	71	19
6-or-more	0	27	130



01



ABOUT ME
AND MY COMPANY

02

RATIONALE
AND RESEARCH QUESTION

03

THEORIES, EMPIRICAL
STUDIES AND GAPS

04

WHAT I'VE DONE

05

WHAT I'VE FOUND

06

DISCUSSION

07

RECOMMENDATIONS



OBJECTIVE 1

Level of occupancy

Binary detection

Multiclass classification
(balanced)

Multiclass classification
(imbalanced)



OBJECTIVE 2

Algorithm & variable

XGBoost -> Hybrid model

Indoor CO2

Ventilation modes -> Ventilation rates

Contextual and periodical validation data

**Other indoor, outdoor air variables:
invalid**



OBJECTIVE 3

Requirements

Periodical validation data
Ventilation rate measurements





01



ABOUT ME
AND MY COMPANY

02

RATIONALE
AND RESEARCH QUESTION

03

THEORIES, EMPIRICAL
STUDIES AND GAPS

04

WHAT I'VE DONE

05

WHAT I'VE FOUND

06

DISCUSSION

07

RECOMMENDATIONS



RECOMMENDATIONS

Strategic approaches	Tactical solutions	Researchers	sustainability.io	Facility team
Validation data	To manually collect occupancy data periodically	x		
	To deploy non-intrusive sensors: IR, heat, door, outdoor air	x	x	x
Ventilation airflow rate measurement	To interview building services engineer and facility staff	x		
	To complete programming code to detect ventilation in real time	x	x	
	To install anemometer in ducts and pipes		x	x
Occupancy-based application deployment for HVAC decisions		x	x	x

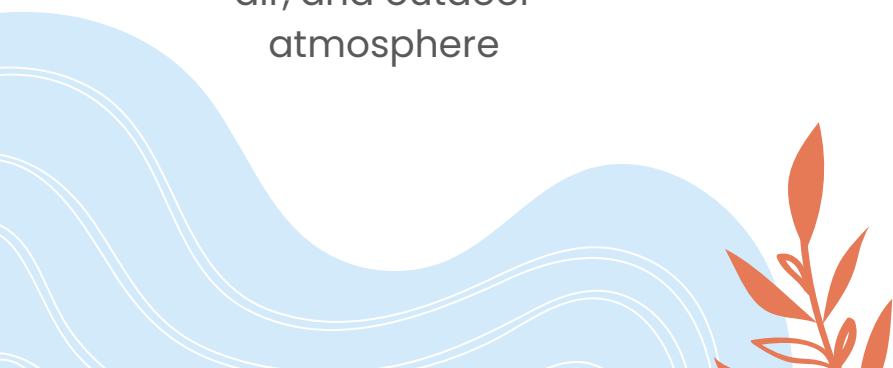


CONTRIBUTION

2-week

data

2-week data of
occupancy, indoor
air, and outdoor
atmosphere



HVAC

room setting

ventilation mode
feature engineering



ML

occupancy

detection and
classification

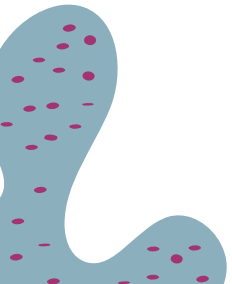
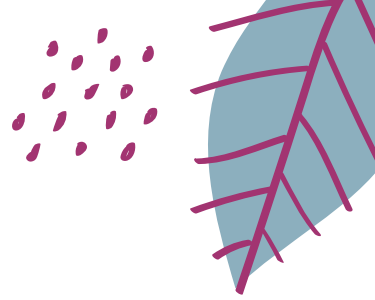


LIMITATIONS

GAPS IN DATASET

UNCERTAINTY

ONLY 1 ROOM





LIMITATIONS & FUTURE RESEARCH

GAPS IN DATASET

UNCERTAINTY

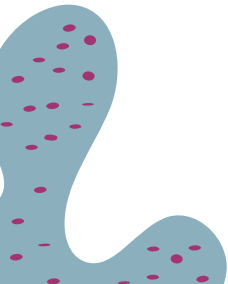
ONLY 1 ROOM

COLLABORATION

WIND SENSOR

DIFFERENT CLASSROOMS

**OCCUPANCY-BASED
APPLICATION**





THANK YOU!

It's time for Q&A.