DAD

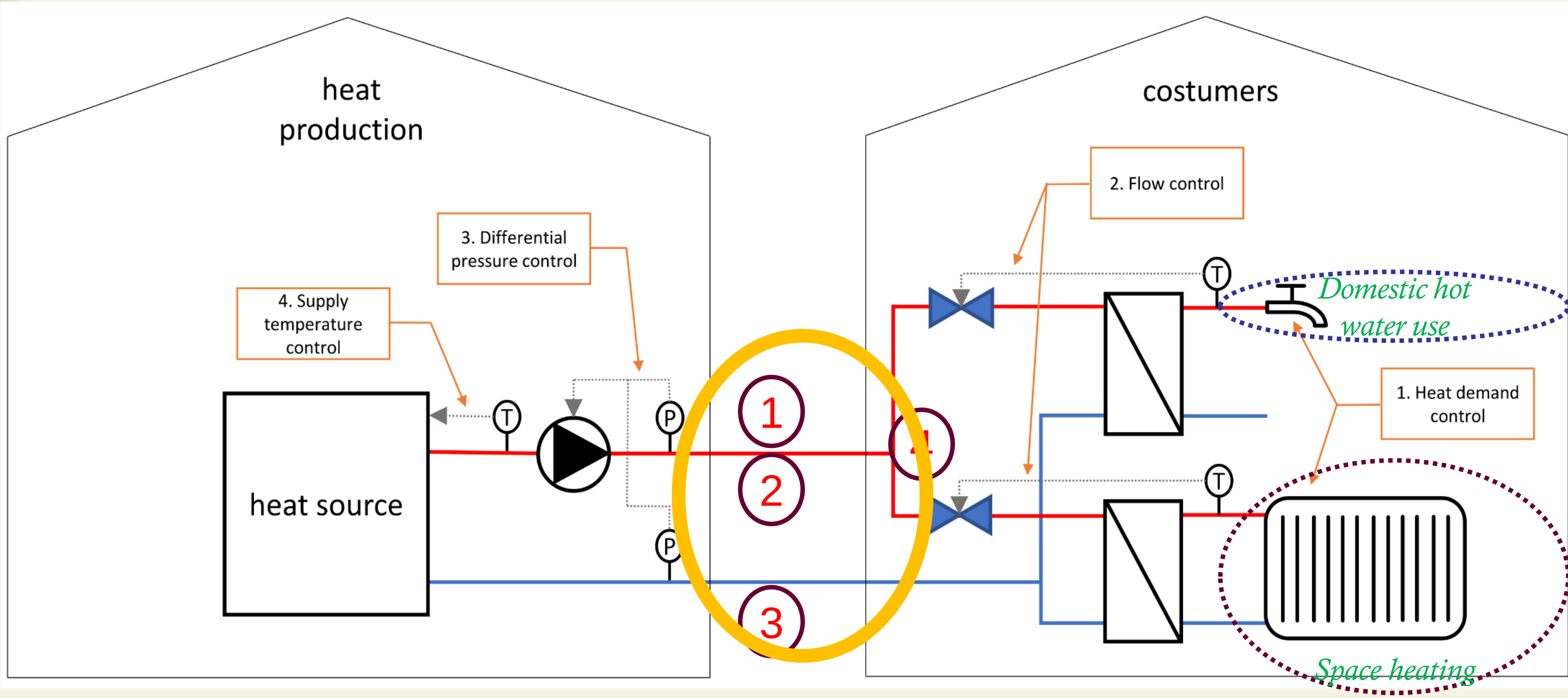
Data Analytics for fault Detection
in district heating systems

How can analytics models be used to detect defects in DH substations?

gideon.mbiydzenyuy@hb.se

District Heating Substation

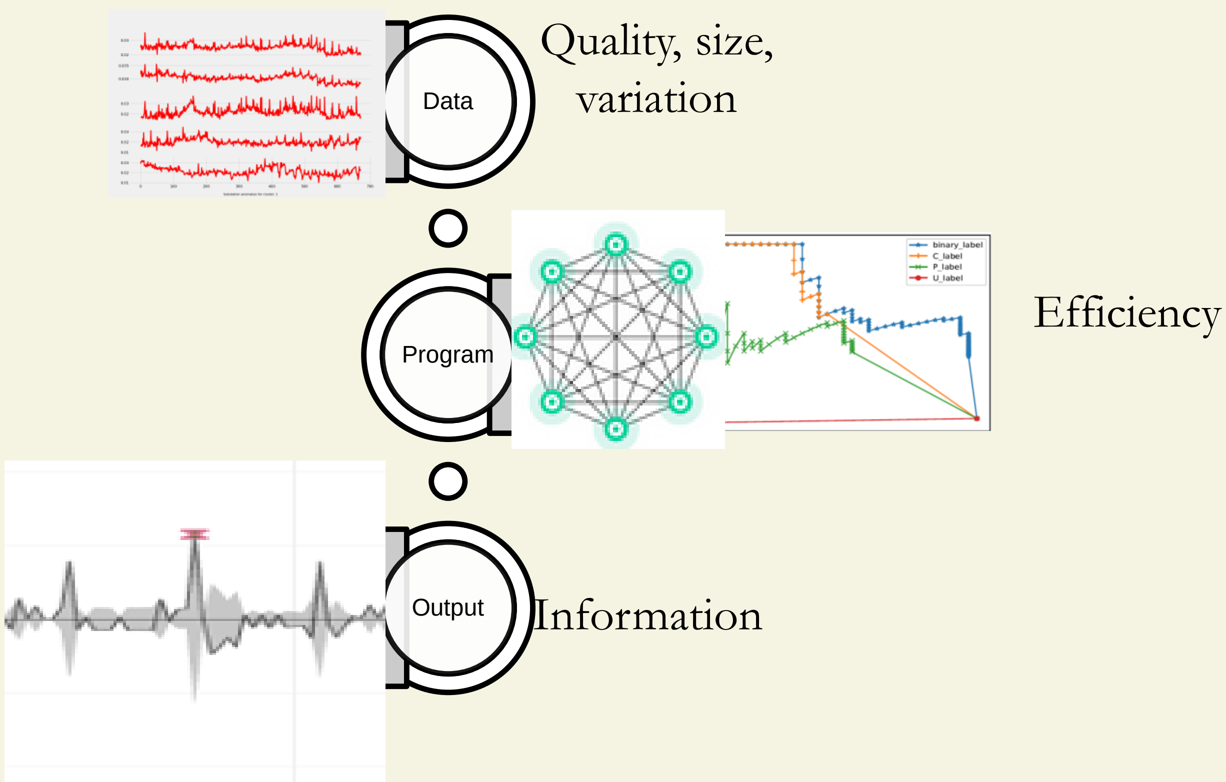
1 *Supply hot water temp in °C*



2 *Hot water volume on arrival in m3* 3 *Return hot water temp in °C* 4 *Energy in MWh*

Data Analytics

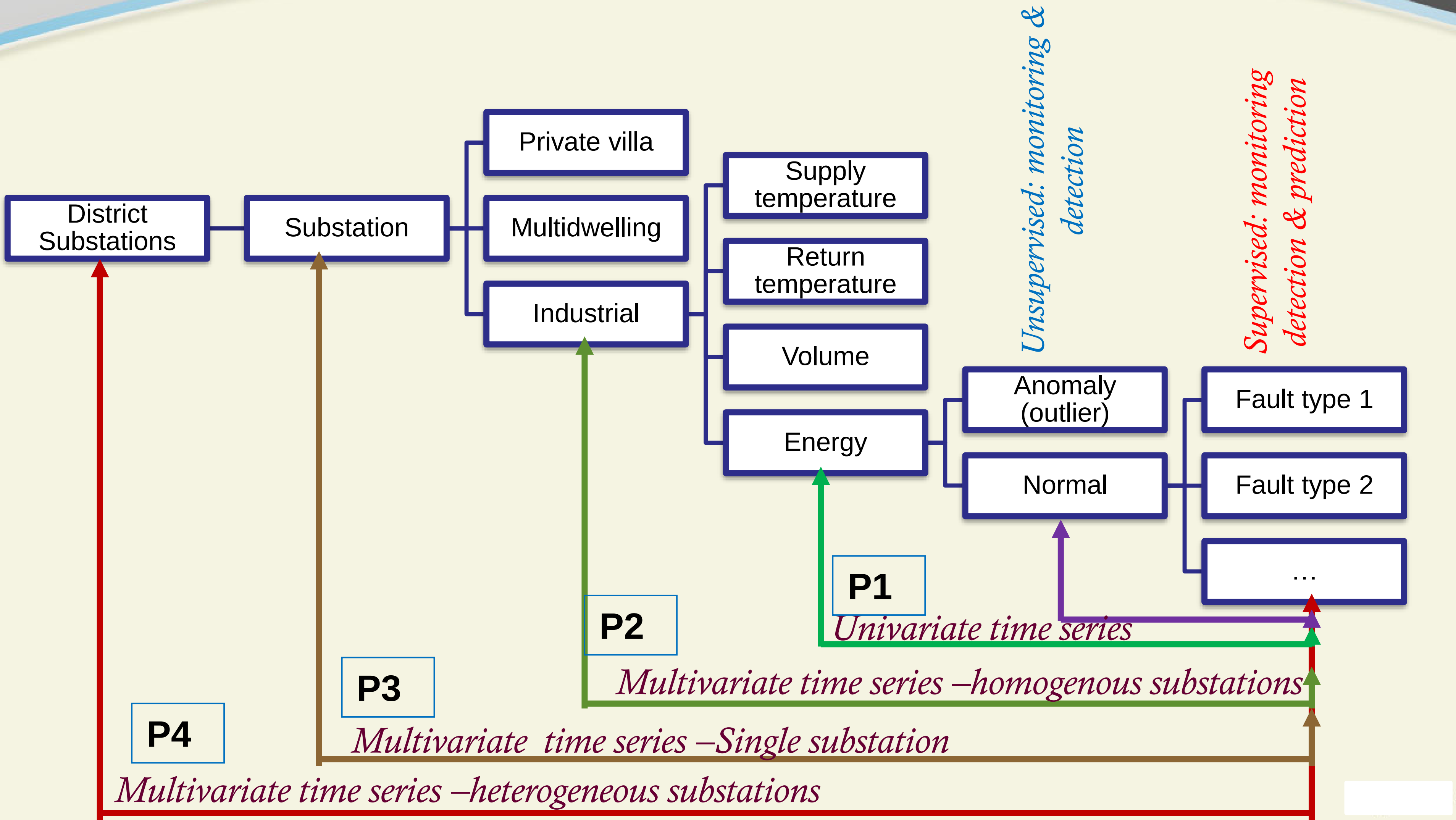
- Machine Learning (ML)



DAD Challenge

Develop ML models to monitor, detect and predict faults in DH substations

Problem context

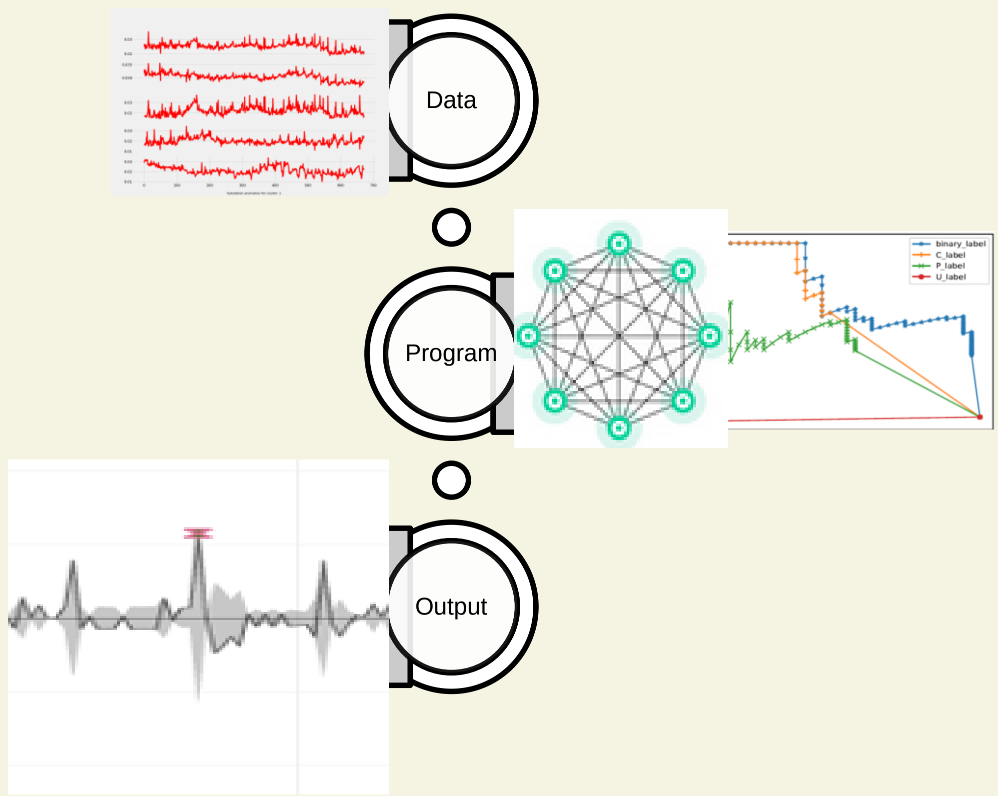


Objectives

- Algorithms:
  - learning deep representations from time series
  - fault detection and prediction with confidence in time series
  - knowledge representation to support real time decision making
- Reduce instability during model learning

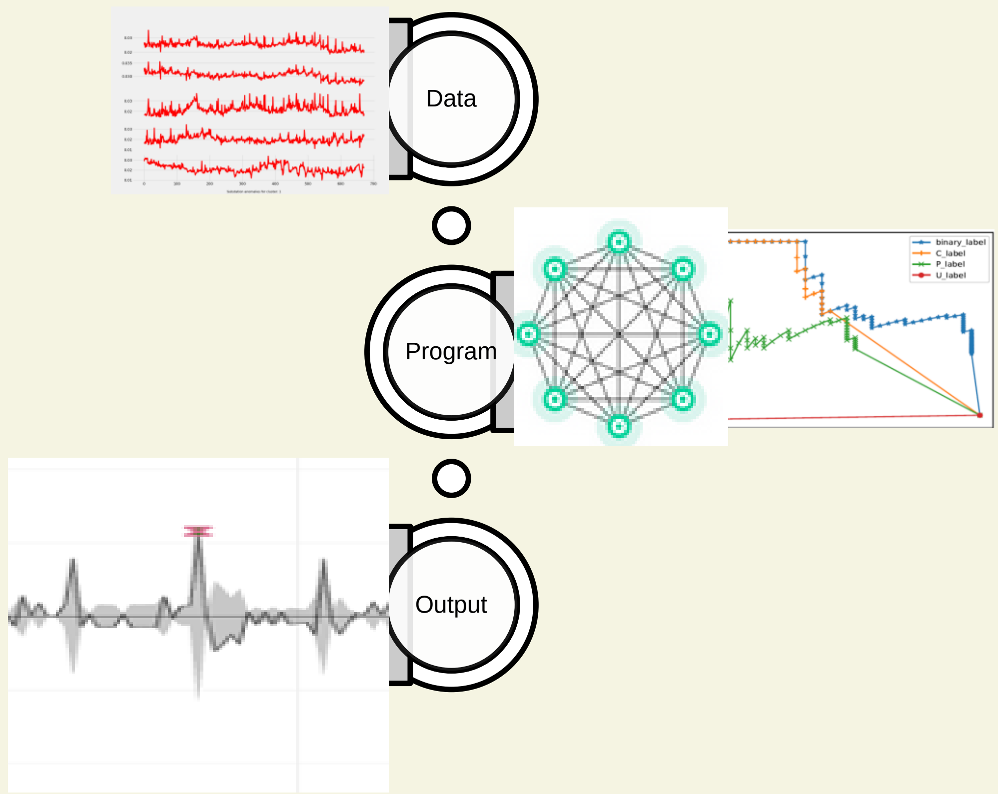
HDLS dataset anomaly detection

- Dataset 2:
  - [2017-01-01   2021-05-18]
  - [4766 rows x 38376 columns]
- Dataset 1:
  - [2017-01-01 2021-03-31]
  - [4766 rows x 10824 columns]



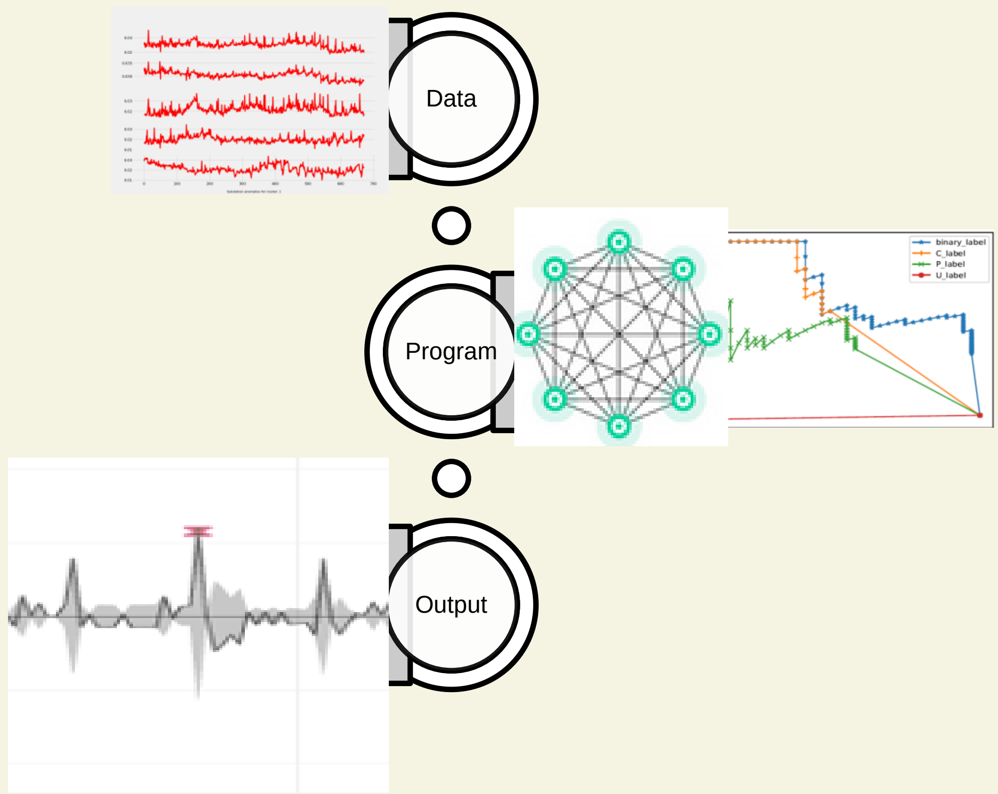
HDLS dataset anomaly detection

- Dataset 2:
  - [2017-01-01   2021-05-18]
  - [4766 rows x 38376 columns]
- Dataset 1:
  - [2017-01-01 2021-03-31]
  - [4766 rows x 10824 columns]
- IQR/MAD
- PCA
- Spectral clustering
- OPTICS-Ordering points to identify the clustering structure
- LSTM Autoencoding
- Rule-based validation model



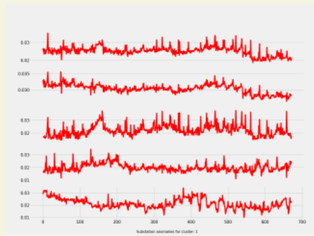
HDLS dataset anomaly detection

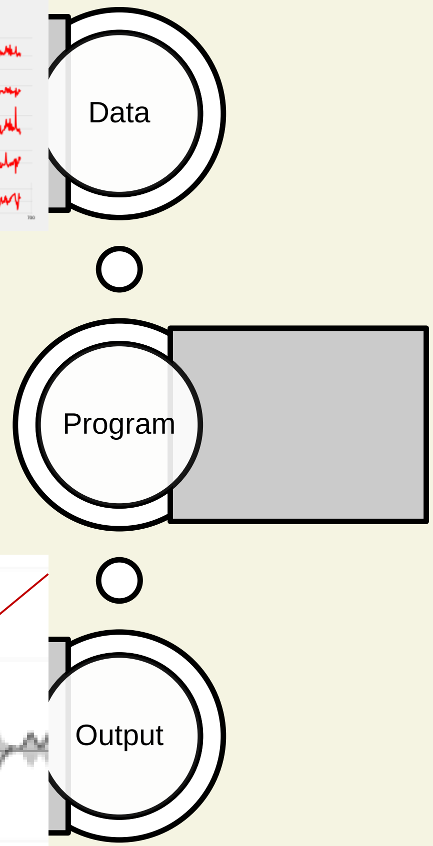
- Dataset 2:
  - [2017-01-01   2021-05-18]
  - [4766 rows x 38376 columns]
- Dataset 1:
  - [2017-01-01 2021-03-31]
  - [4766 rows x 10824 columns]
- **IQR/MAD**
- **PCA**
- **Spectral clustering (SC)**
- **OPTICS-Ordering points to identify the clustering structure**
- LSTM Autoencoding
- Rule-based validation model



PCA, SC, OPTICS

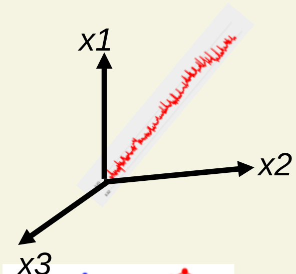
- Find k groups of most similar substations

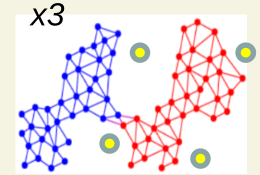


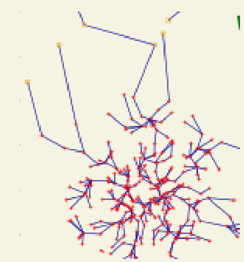


PCA+MAD

SC + OPTICS







Why do this?

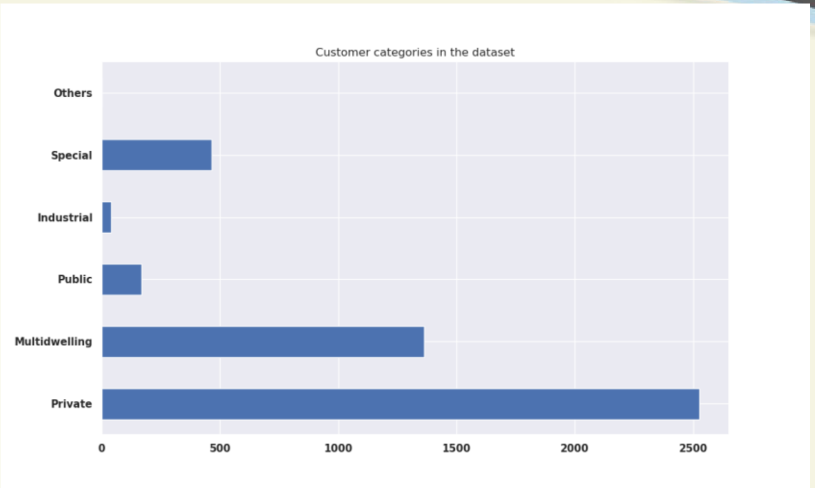
- The nature of the data:
  - Highly dimensional dataset
  - Large number of substations
  - No labels
  - Low quality
  - Claim: If an energy profile is too different from hundreds or thousands of energy profiles in the same DH network, it probably has an issue!

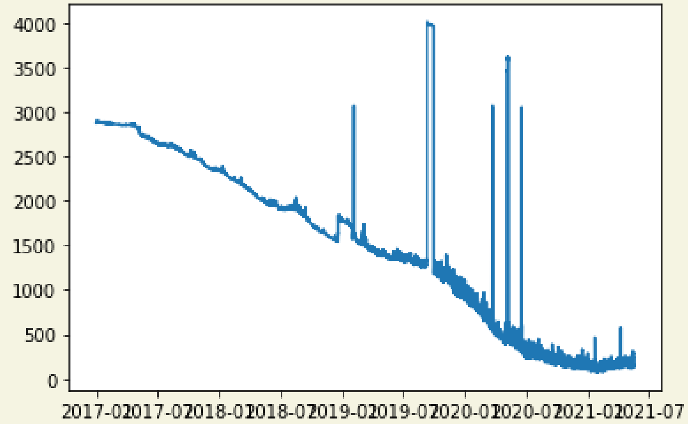
- Claim:
  - If a substation energy profile is uniquely different from hundreds or thousands of energy profiles in the same DH network, then it probably has an issue that deserve attention!

- Counter claim:
  - If a substation energy profile is uniquely different from hundreds or thousands of energy profiles in the same DH network, then there is no issue that deserve attention if we can satisfactorily explain it uniqueness.

Limitations

- PCA – how many factors to consider
- SC - which value of k?
- OPTICS – how should density points be connected
- Stable results
- Variations in distributions of the data
- What can we learn from an anomaly?





Experiment

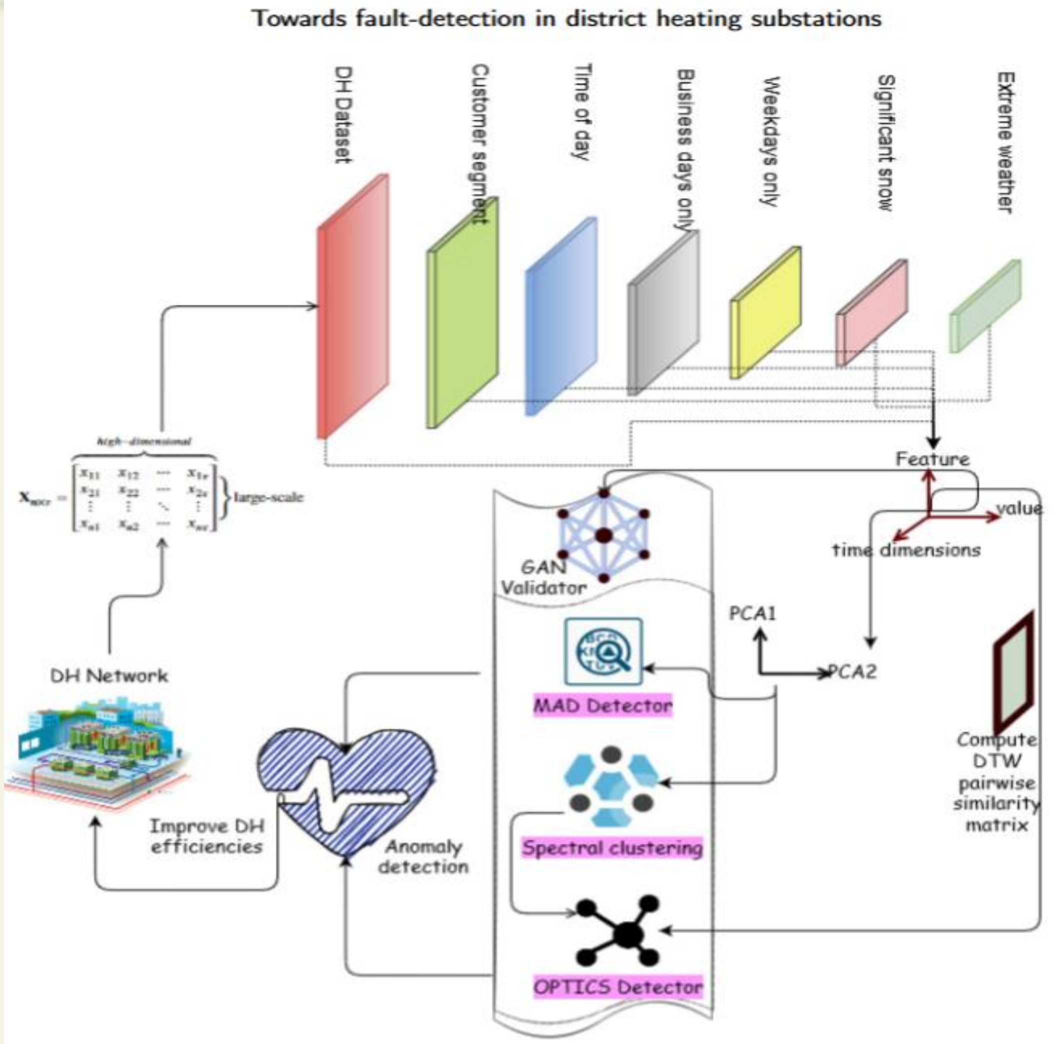
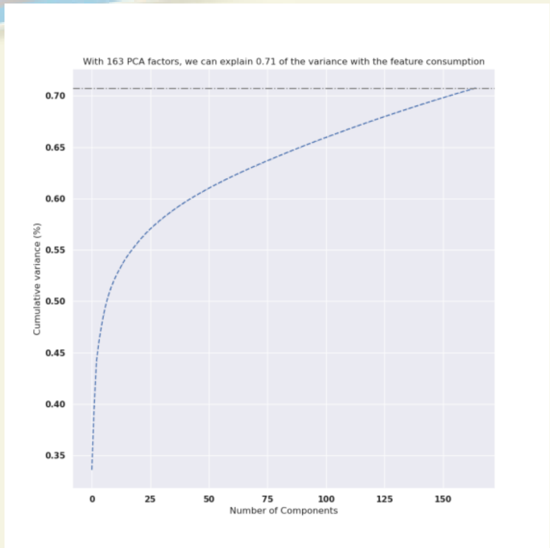
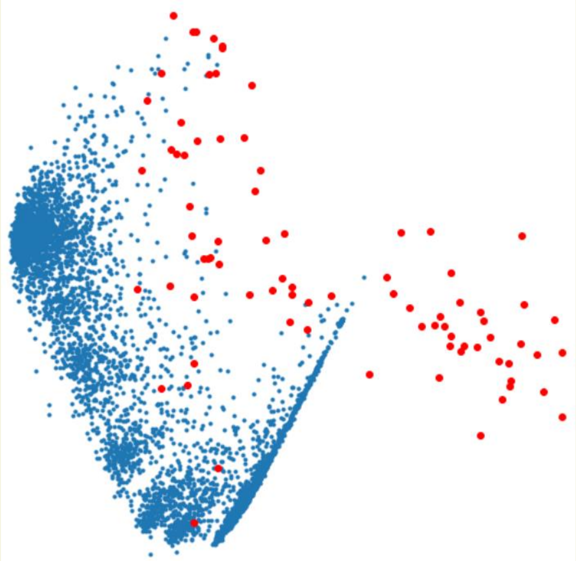
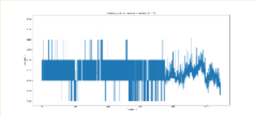


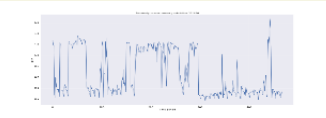
Figure 1: Method overview.

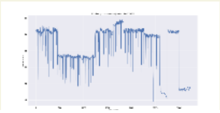
PCA+MAD

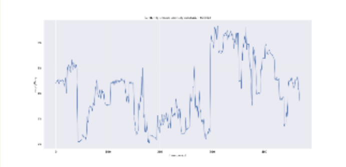

With 163 PCA factors, we can explain 0.71 of the variance with the feature consumption
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
Cumulative variance (%)
0
25
50
75
100
125
150
Number of Components

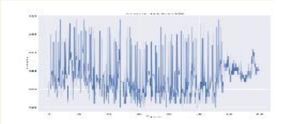


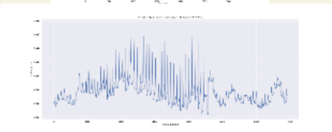


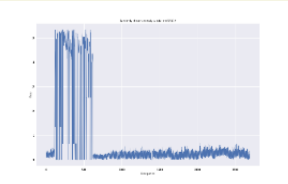


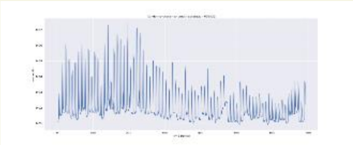




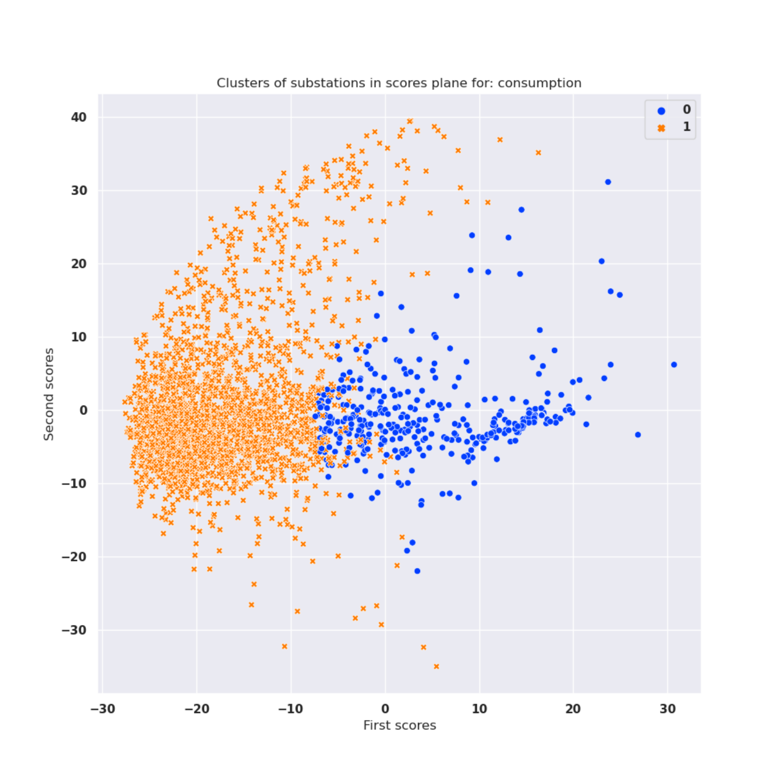


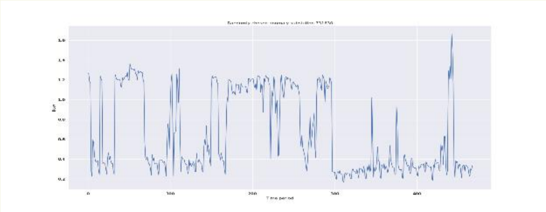


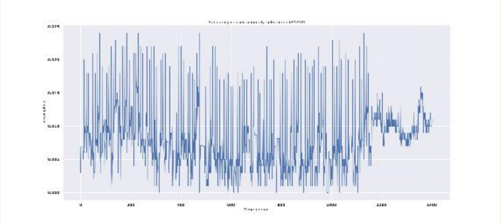


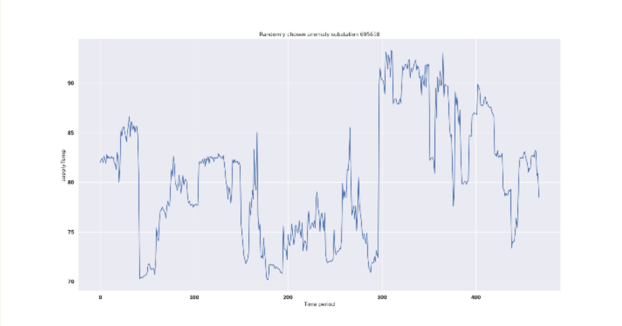


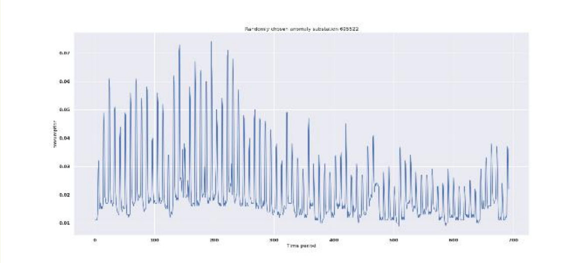
SC+OPTICS


Clusters of substations in scores plane for: consumption
0
1
40
30
20
10
0
−10
−20
−30
Second scores
−30
−20
−10
0
10
20
30
First scores







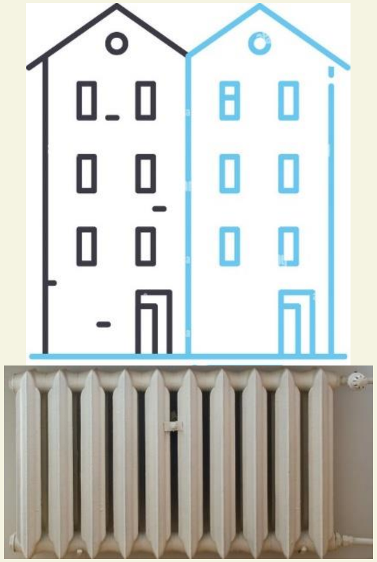


Data Driven Insights

- How does the system behavior changes in varying operational conditions?
  - Scenario **1-space heating** **[22:00 – 05:00]** energy consumption in **multidwelling** customer facilities
  - Scenario 2 -**space heating and domestic hot water** use **[05:00 – 22:00]** in **multidwelling** customer facilities
  - Scenario 3 -Scenario 3, **space heating** and **domestic hot water** use in **all types of customer facilities**
  - Scenario 4 -**space heating** and **domestic hot water** use in **all types of customer facilities** for all sensor features

Scenario 1

- Target space heating in multidwelling customer facilities alone
  - << 1% anomaly rate across all features
- Open Question: Is the noise in substation data unrelated to space heating itself?

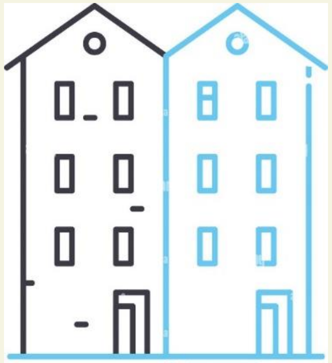


Scenario 2

- Target domestic hot water and space heating in multidwelling customer facilities
  - ~4% anomaly rate
  - Between 4 to 8% for all features

- Open Question: Is hot water use the source of unusual behavior in multidwelling customer facilities?

- Unaccounted heat activities: ventilation, use of doors and windows as well as micro heat sources including cooking, electronic appliances and air pumps

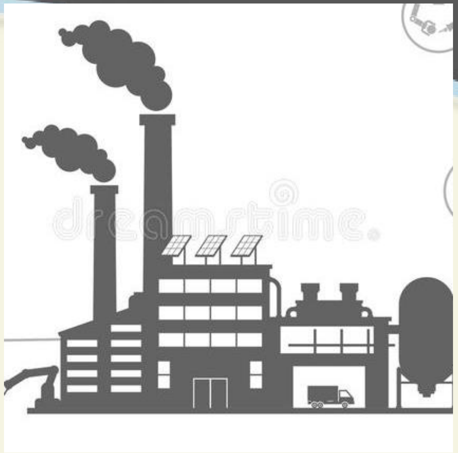


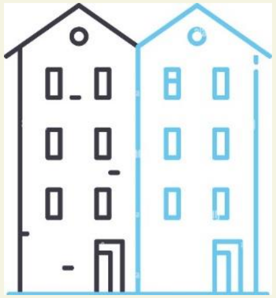
Scenario 3

- Target domestic hot water and space heating in all types of customer facilities (consumption in MWh alone)
  - ~12% anomaly rate (combining rate from each type of customer facility)
  - ~23% anomaly rate (all facilities together)

- Expected: increase data variability → increase of possibility of abnormal energy signatures

- Open Question: Is it a good idea to combine customer facilities when performing anomaly detection on a DH network level?



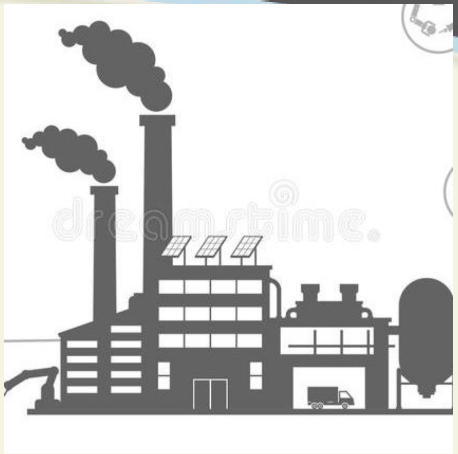


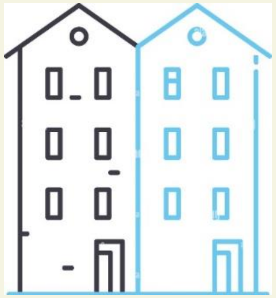
Scenario 4

- Target domestic hot water and space heating in all types of customer facilities for all types of sensor features
  - ~43% anomaly rate (combining rate from each feature)

- Comments → Results aligning with values from literature

- Open Question: which sensor characteristics are most effective for anomaly detection in substations and why?



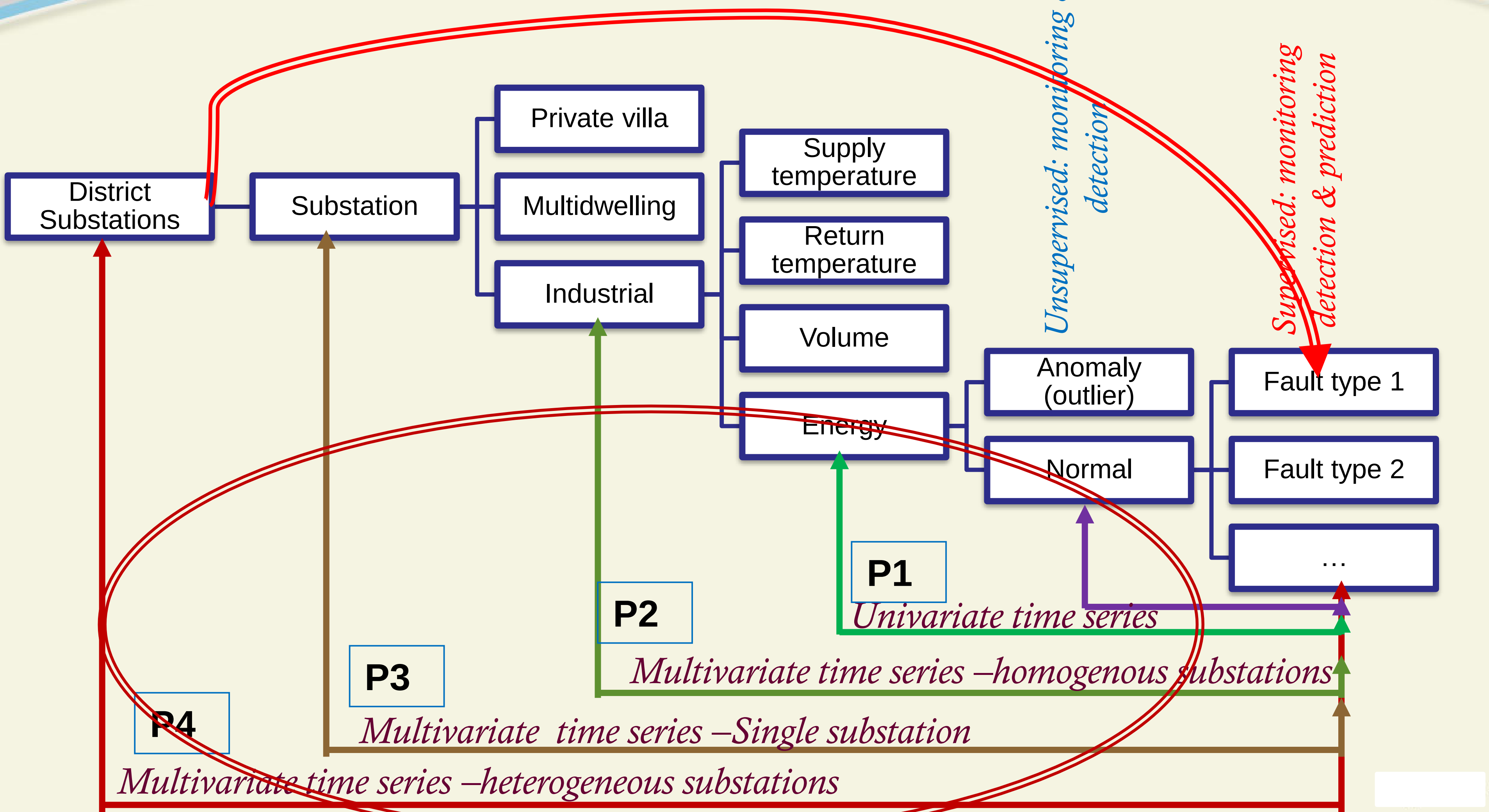


Scenario 5

- How does weekends and holidays affect abnormal energy signals?
  - Without weekends and holidays:
  - ~34% anomaly rate
  - With weekends and holidays:
  - ~28% anomaly rate
  - With snow and extreme weather:
  - ~39% anomaly rate

- Open Question: how should we handle weekends and holidays when training models to be deployed in monitoring networks?

Key message



District Substations
Substation
Private villa
Multidwelling
Industrial
Supply temperature
Return temperature
Volume
Energy
Unsupervised: monitoring & detection
Supervised: monitoring detection & prediction
Anomaly (outlier)
Normal
Fault type 1
Fault type 2
…
P1
Univariate time series
P2
Multivariate time series –homogenous substations
P3
Multivariate time series –Single substation
P4
Multivariate time series –heterogeneous substations

DAD Partners

KK·stiftelsen

DAD is a KK-HÖG 2017 initiative
Funding: 50% ca 3.8 MKr
Duration: 2018-01-01 to 2020-12-31.

Project home page:
https://www.hb.se/en/Research/Research-Portal/Projects/Data-Analytics-for-Fault-Detection-in-District-Heating-DAD/

UNIVERSITY OF BORÅS

NODA Intelligent Systems

BORÅS ENERGI OCH MILJÖ

AB BOSTÄDER I BORÅS

DAD Team

KK·stiftelsen

UNIVERSITY OF BORÅS

patrick.gabrielsson@hb.se

gideon.mbiydzenyuy@hb.se

anders.gidenstam@hb.se

henrik.linusson@hb.se

ulf.johansson@hb.se

NODA Intelligent Systems

BORÅS ENERGI OCH MILJÖ

AB BOSTÄDER I BORÅS

cj@noda.se

Andreas.Carlsson

josefine.norresjo@bo

Unsupervised detection of anomalies in high-dimensional large-scale data using Clustering★

Gideon Mbiydzenyuy[a,*,1], Federico Benzi[b], Anders Gidenstam[c], Jens Brage[d], Håkan Sundell[e] and Ulf Johansson[f]

[a] *University of Borås, Allégatan 1, Borås, SE-501 90, Sweden*
[a] *SOLITA, Tulegatan 11, Stockholm, SE-113 53, Sweden*
[a] *University of Borås, Allégatan 1, Borås, SE-501 90, Sweden*
[a] *NODA, Biblioteksgatan 4, Karlshamn, SE-374 35, Sweden*
[a] *University of Borås, Allégatan 1, Borås, SE-501 90, Sweden*

CSL@BS

Forskningsinriktningar är maskininlärning, data mining, high performance computing och distributed computing.
https://www.hb.se/forskning/forskningsportal/forskargrupper/cslbs---computer-science-lab-at-boras-sweden/


INSTITUTIONEN FÖR INFORMATIONSTEKNOLOGI

Vår forskning

Våra utbildningar

Samverkan

HÖGSKOLAN I BORÅS / OM HÖGSKOLAN / ORGANISATION / AKADEMIER OCH ENHETER / AKADEMIER / AKADEMIN FÖR BIBLIOTEK, INFORMATION, PEDAGOGIK OCH IT / INSTITUTIONEN FÖR INFORMATIONSTEKNOLOGI

Institutionen för informationsteknologi

Hos oss på institutionen bedrivs utbildning och forskning inom ämnet informatik i en miljö med fristående kurser och utbildningsprogram på kandidat- och magisternivå.

https://www.hb.se/om-hogskolan/organisation/akademier--enheter/akademier/akademin-for-bibliotek-information-pedagogik-och-it/institutionen-for-informationsteknologi/