Dwelling Energy Insights – Week 7

Team members: Merel Kreszner, Olivier van Luijk, Rajeev Kalloe,
Santiago Puertas Puchol, Emilio Caba Batuecas & Teo Čurčić

Product Owner: Dr. ir. T.B. Salcedo Rahola

Main Research Question

- The main research question:
- How can we predict dwelling groups based on their similarities in energy usage and behavior?

- Sub-questions:
- What is the maximum/minimum/average usage/production per day per household?
- What is the maximum amount of Ampere that a household delivers?
- What kind of classification can we produce based on the dwelling characteristics?

Answer of Question 2

	max_prod_date	max_prod	max_cons_date	max_cons
H01	2018-07-02	32.651	2018-03-01	59.567
H02	2018-06-26	22.968	2017-12-31	61.718
H03	2018-07-05	15.145	2018-03-03	25.126
H04	2018-06-26	20.181	2017-12-29	21.525
H05	2018-06-13	19.169	2019-01-21	19.637
H06	2019-05-14	22.617	2018-01-13	40.147
H07	2018-06-26	24.840	2018-03-01	75.839
H08	2018-06-26	14.688	2017-12-31	66.913
H09	2018-07-03	26.461	2017-12-29	60.675
H10	2019-05-24	22.049	2018-12-15	93.523

max_prod_amperes_date	max_prod_amperes	max_cons_amperes_date	max_cons_amperes
2018-07-02 13:45:00	15.843478	2017-12-11 18:00:00	36.243478
2019-05-05 13:30:00	13.443478	2018-02-17 09:30:00	33.130435
2019-05-12 13:15:00	8.800000	2017-10-03 18:15:00	24.086957
2018-06-21 14:00:00	11.130435	2019-01-20 21:45:00	17.495652
2018-06-21 14:00:00	12.139130	2018-10-20 18:15:00	23.286957
2019-05-12 13:15:00	13.391304	2017-12-13 18:30:00	33.286957
2018-05-21 13:45:00	13.686957	2018-03-01 19:45:00	45.321739
2018-06-21 14:00:00	8.121739	2017-11-25 08:15:00	38.121739
2017-10-12 12:30:00	17.286957	2018-02-24 17:30:00	37.165217
2019-04-13 13:00:00	12.852174	2019-03-16 07:15:00	43.686957

Progress so far

- Made new visualizations of consumption per type of heating system per dwelling
- We finished cleaning our dataset, or did we?
- We started implementing the following Machine Learning Algorithms for classification:
- Logistic Regression
- K-Nearest Neighbours
- Support Vector Machine

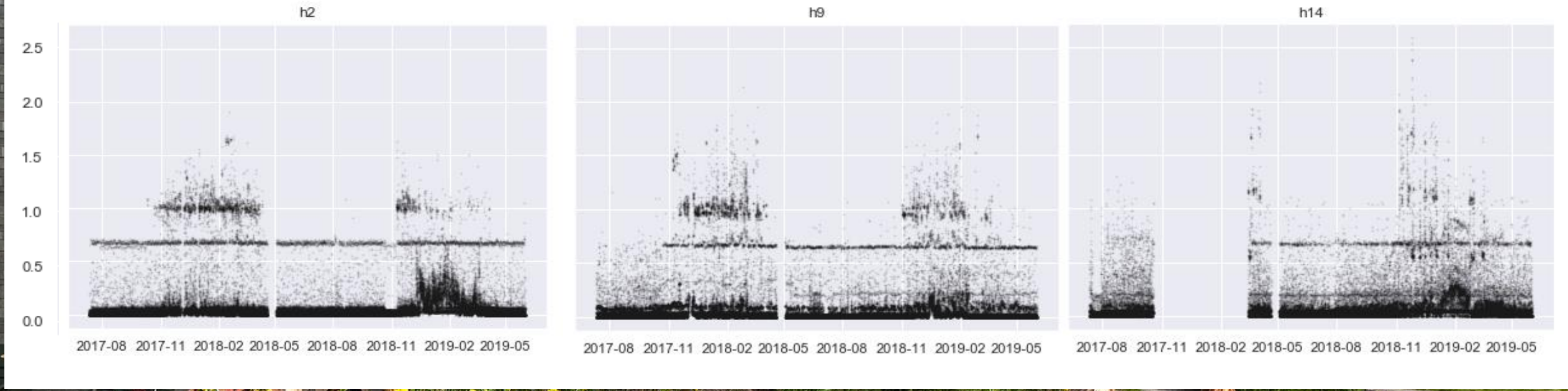
Impediments

- Unfortunately we didn’t clean our data correctly, because all the columns had different lengths, which the algorithms don’t accept.

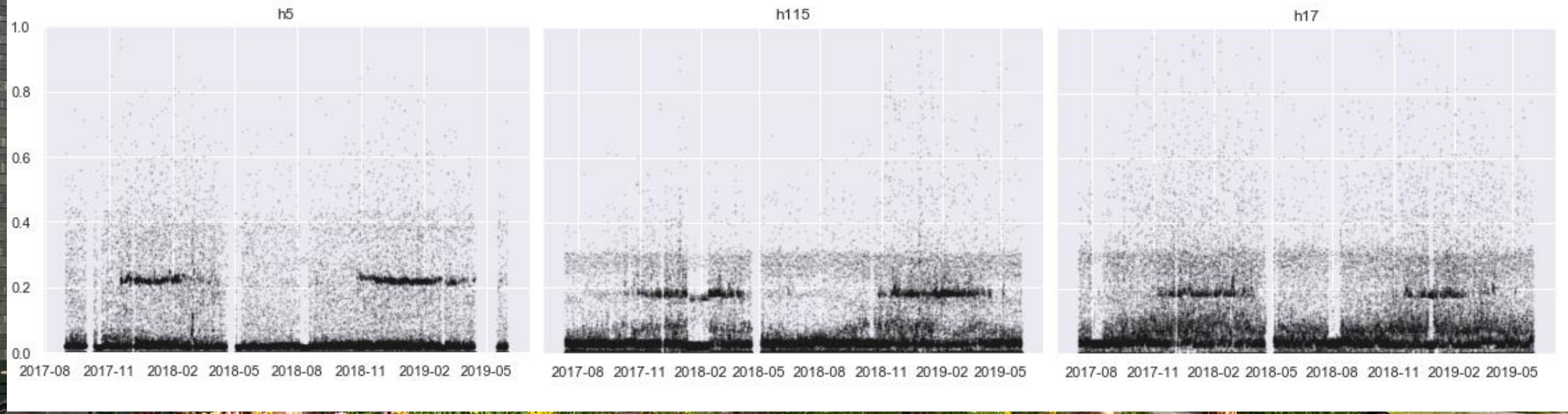
- New approach in cleaning the data:
  - Looking for outliers/NaN on the whole dataset and remove them.

- Next step:
  - Creating dummy variables based on the data.

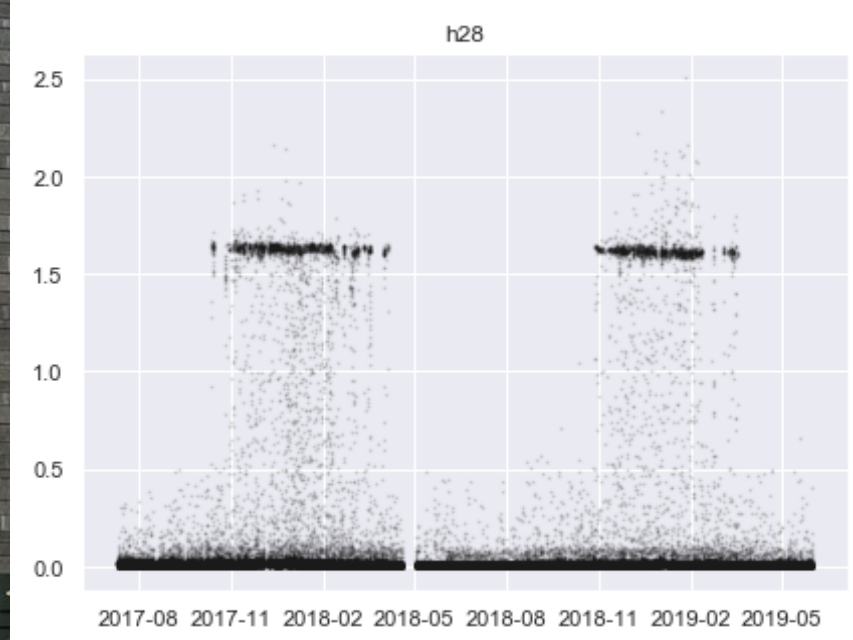
Electricity

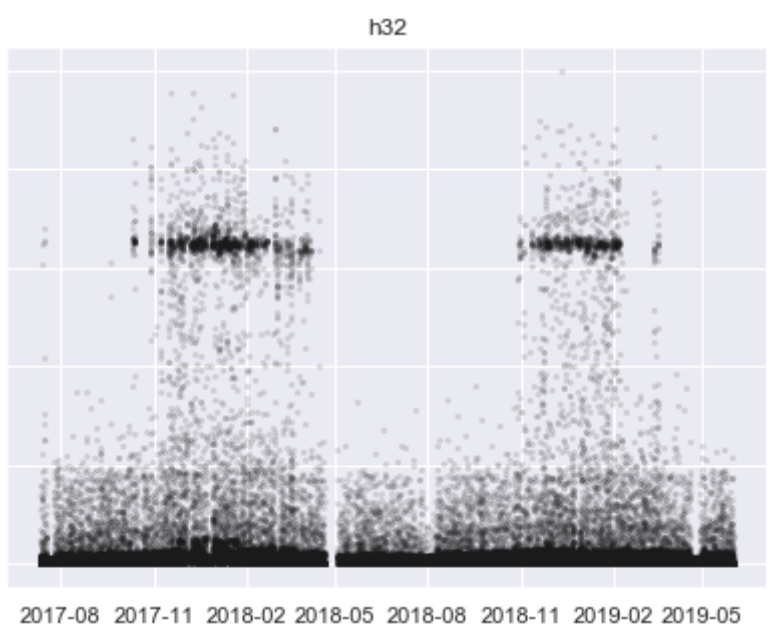


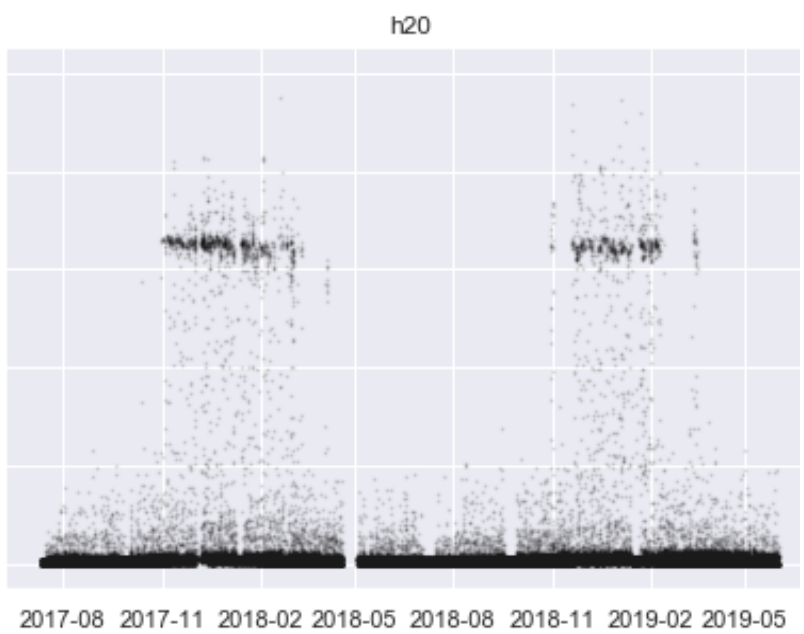
Heat pump


h5
h115
h17
1.0
0.8
0.6
0.4
0.2
0.0
2017-08
2017-11
2018-02
2018-05
2018-08
2018-11
2019-02
2019-05

Thermal solar panel







Remaining tasks/issues

- Finish the data cleaning process (with the dummy variables)
- Implement T-distributed Stochastic Neighbor Embedding Algorithm
- Implement different kinds of Machine Learning Algorithms

Questions/Feedback

- Are there any questions or feedback based on this presentation?