Project EXPE-SMARTHOUSE

Power consumption prediction in a smart house based on high power consuming devices

Sander Thibau 3A-ASI 2022-2023

Table of contents

Abstract

Machine learning algorithms can be trained on historical power consumption data to make accurate predictions of future power consumption, which can inform energy planning and optimization decisions to promote sustainability and reduce energy waste. The goal of this project is to predict the general power consumption for the next half hour of a smarthouse based on the current power consumption of high consuming devices. K-Nearest Neighbors (k-NN) and Decision Trees are the used machine learning algorithms for building two predictive models. These models are evaluated and compared. The results of this project are discussed regarding its limitations and usefulness.

Introduction

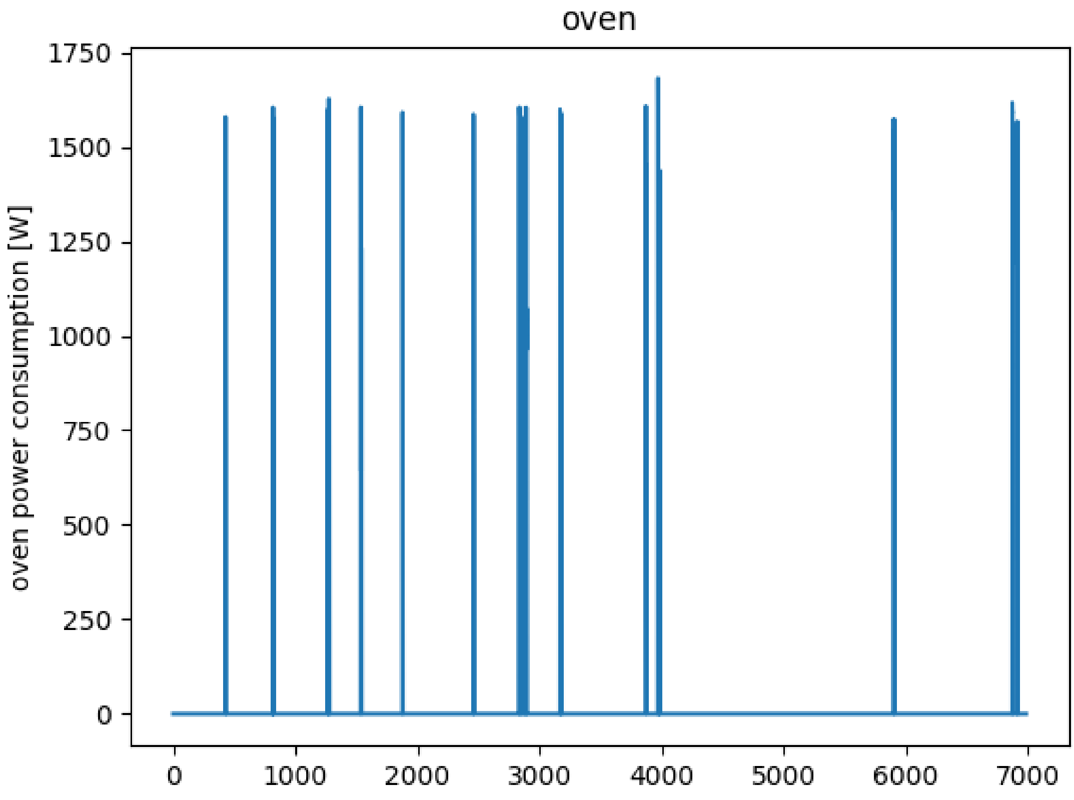
Power consumption prediction is a crucial aspect in modern society as it helps in managing energy resources efficiently. Accurate predictions of power consumption can assist energy providers in optimizing their energy generation and distribution strategies, reducing energy waste and reducing the carbon footprint. Additionally, it can also help businesses and households to plan their energy usage and make informed decisions regarding energy conservation. Power consumption prediction plays a vital role in promoting sustainability and reducing the impact of human activities on the environment.

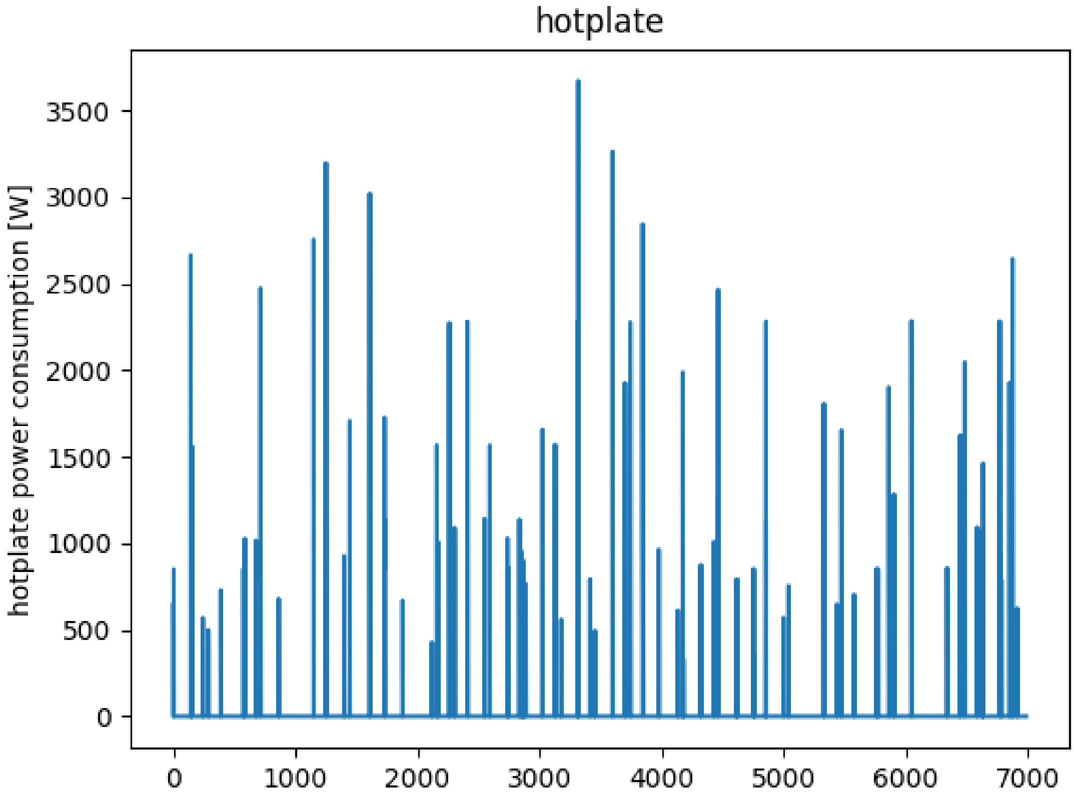
Machine learning can be used for power consumption prediction by training algorithms on historical data to build predictive models. The models can then be used to make predictions of future power consumption based on various factors such as weather, time of day, day of the week, and historical consumption patterns.

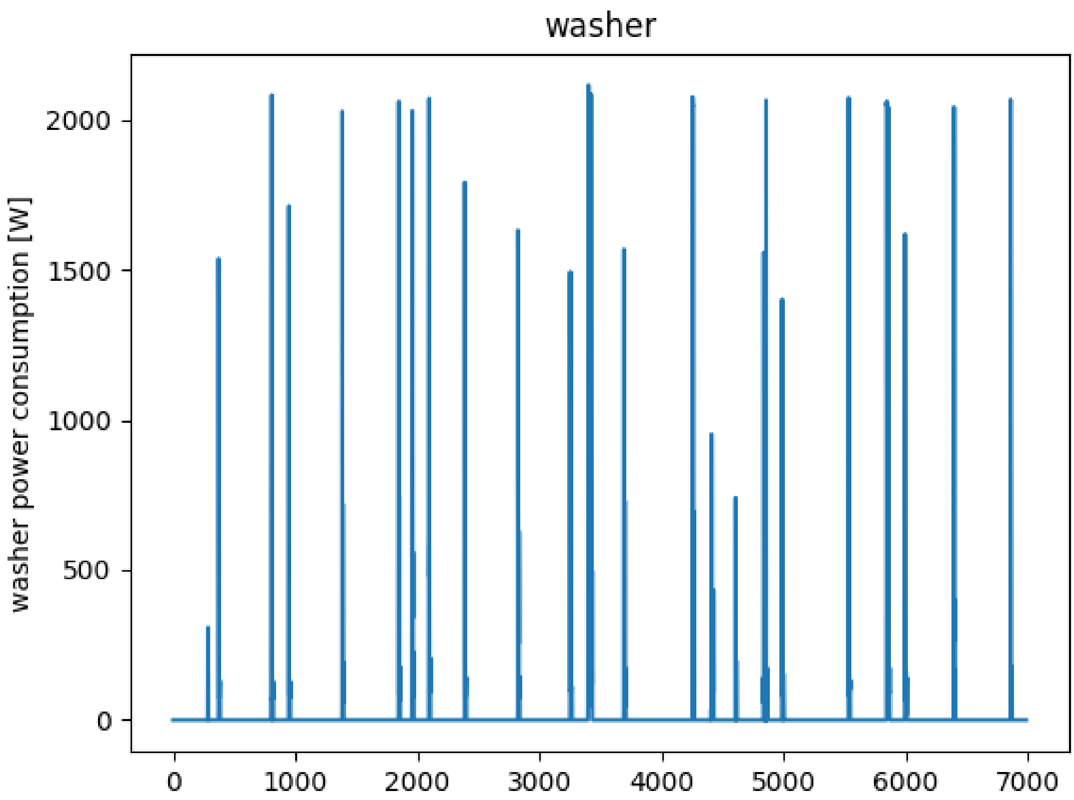
Once the machine learning model is trained and validated, it can be used to make predictions of future power consumption with high accuracy. These predictions can then be used to inform energy planning and optimization decisions, helping to reduce energy waste and promote sustainability.

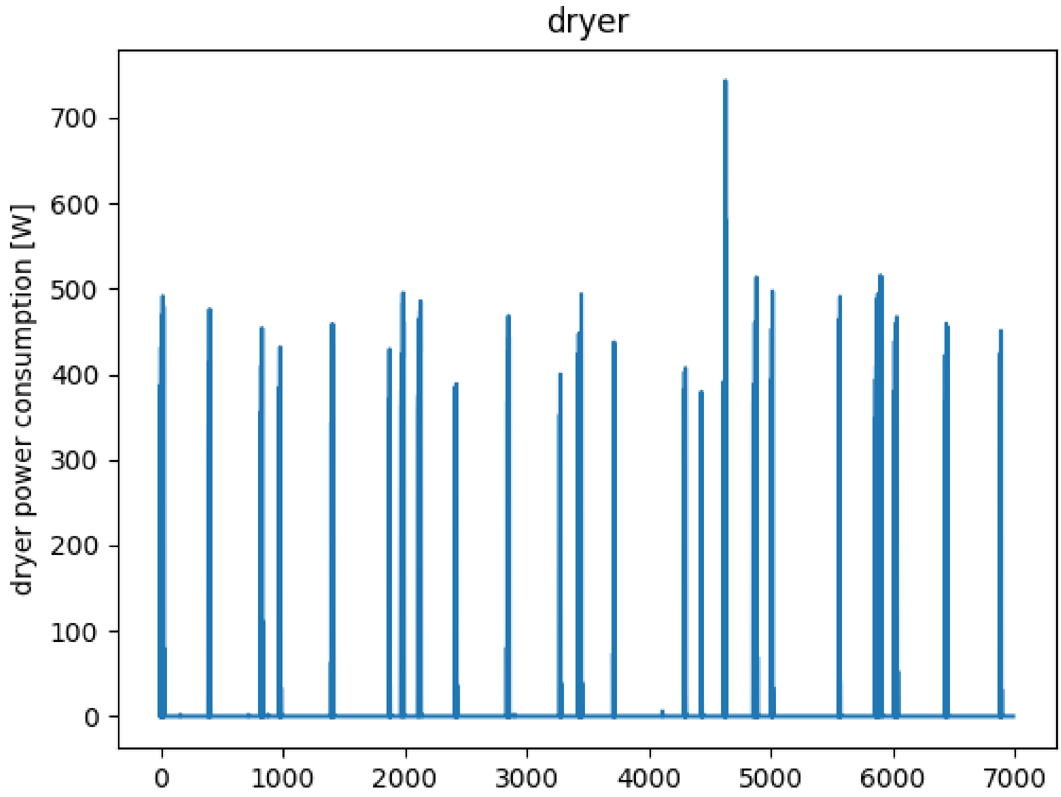
Data cleaning

The data is retrieved from the expe-smarthouse project. The used data is the power consumption of the oven, the hotplate, the washer, the dryer and the total power consumption from 27 November 2022 to 29 January 2023. Between these dates, there was a period with no data. This is excluded from the dataset in further processing. After retrieving the raw data of the server, the dataset is sampled and filled with zeros where there is no data in order to have values every 10 minutes. All data is merged in one csv-file and treated using the pandas module.

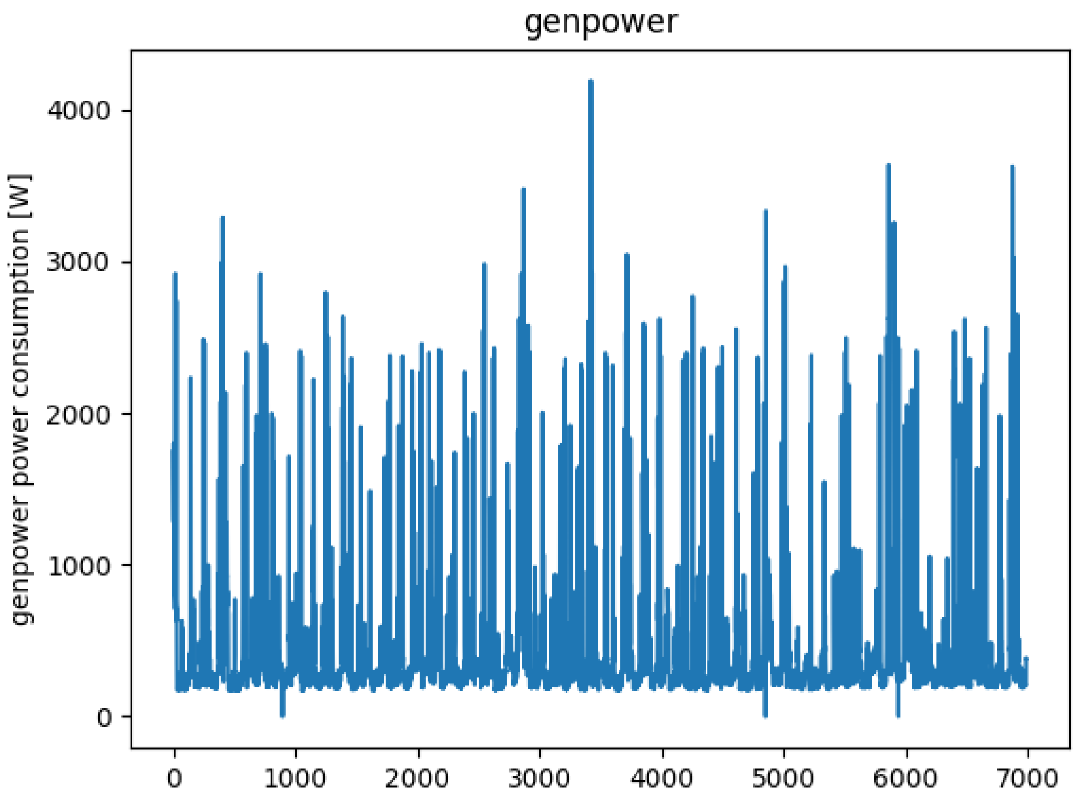


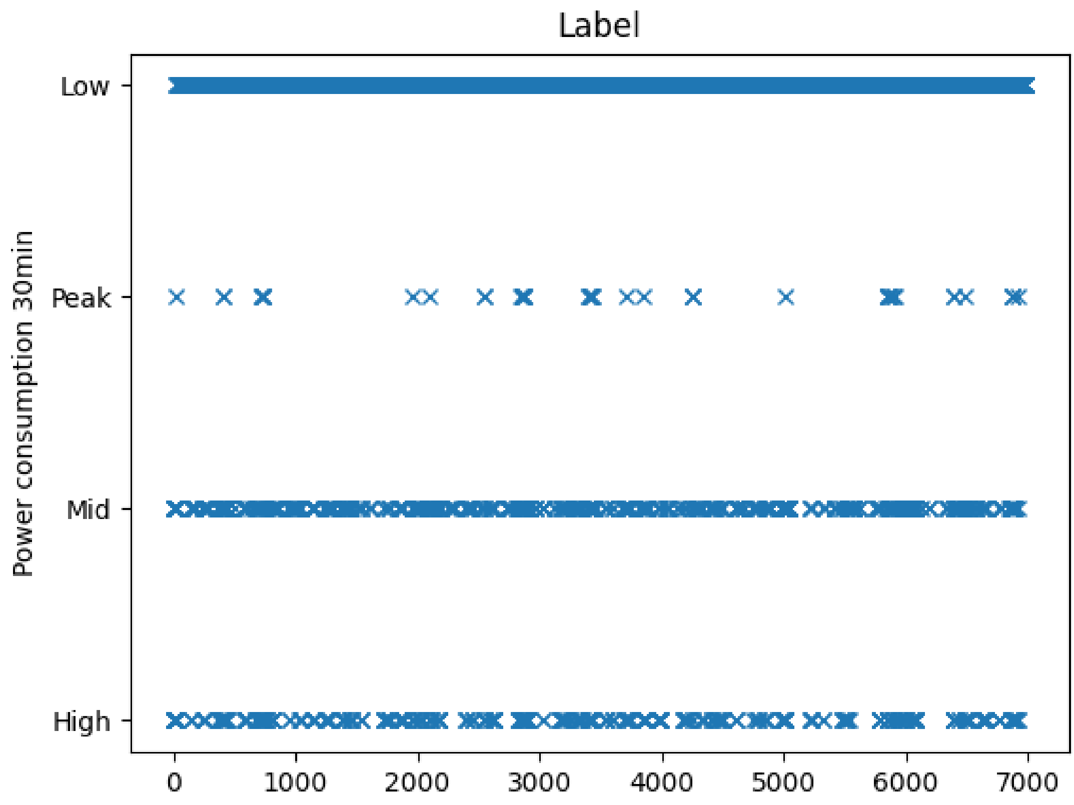

hotplate
hotplate power consumption [W]
3500
3000
2500
2000
1500
1000
500
0
0
1000
2000
3000
4000
5000
6000
7000


washer
washer power consumption [W]
2000
1500
1000
500
0
0
1000
2000
3000
4000
5000
6000
7000



With this information the goal is to predict the average general power consumption of the house for the next half hour. The label is calculated based on the total power consumption, but the prediction is done with the data at one point of time and the label is calculated from the following moments. The label at every timestamp is appended by adding the three following values of the general power consumption together and divided by 3. This is considered the true average general power consumption of the next half hour. The label is then separated in four classes. Class 1 is for values from 0 to 500W and indicated as 'Low' (power consumption), class 2 is for values between 500W and 1000W and indicated as 'Mid', class 3 is for values between 1000W and 2000W and indicated as 'High', class 4 is for values higher than 2000W and indicated as 'Peak'.


genpower
genpower power consumption [W]
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000
6000
7000


Label
Power consumption 30min
Low
Peak
Mid
High
0
1000
2000
3000
4000
5000
6000
7000

k-NN

In a first attempt to predict the power consumption on this limited dataset the k-Nearest Neighbors algorithm is implemented using python sklearn KNeighborsClassifier. To include the influence of the time of the day in the prediction, the hour of the datapoints is appended as a feature. Mainly because of this, the data needs to be rescaled as the algorithm is based on distance. The dataset is divided in 80 percent training data and 20 percent test data. Cross validation is done to find an optimal amount of neighbors and results in k equal to 50.

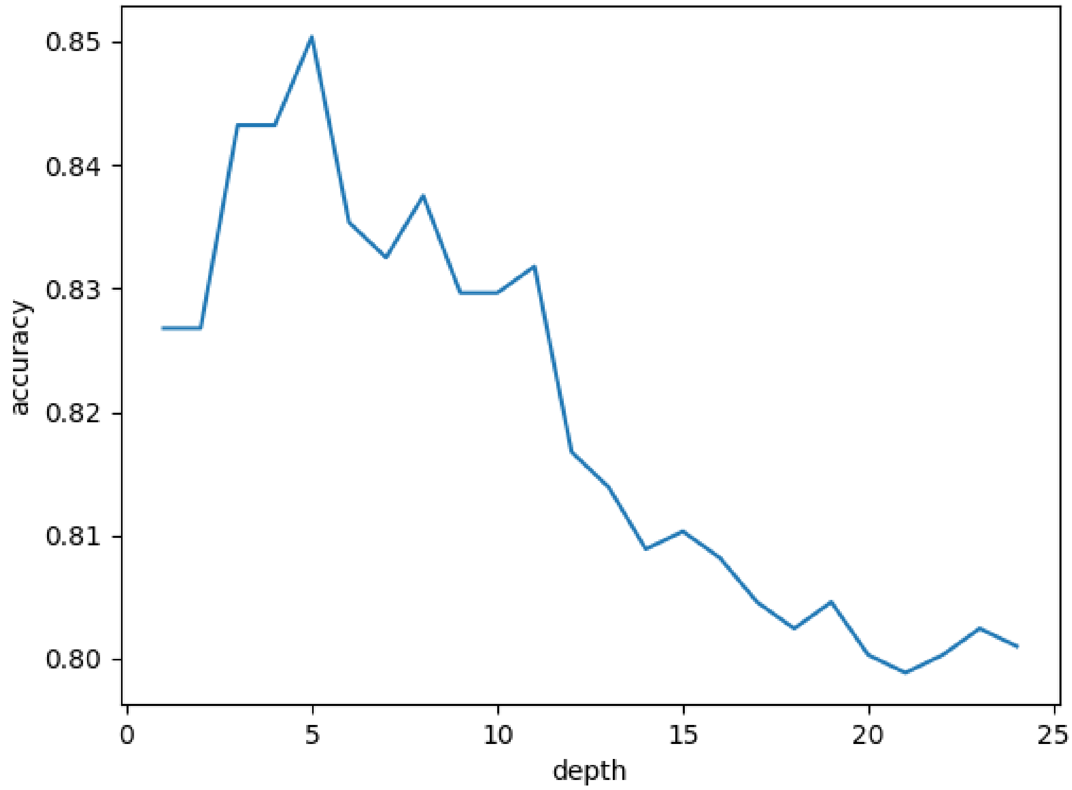
In table 1, the metrics of the classification can be found.

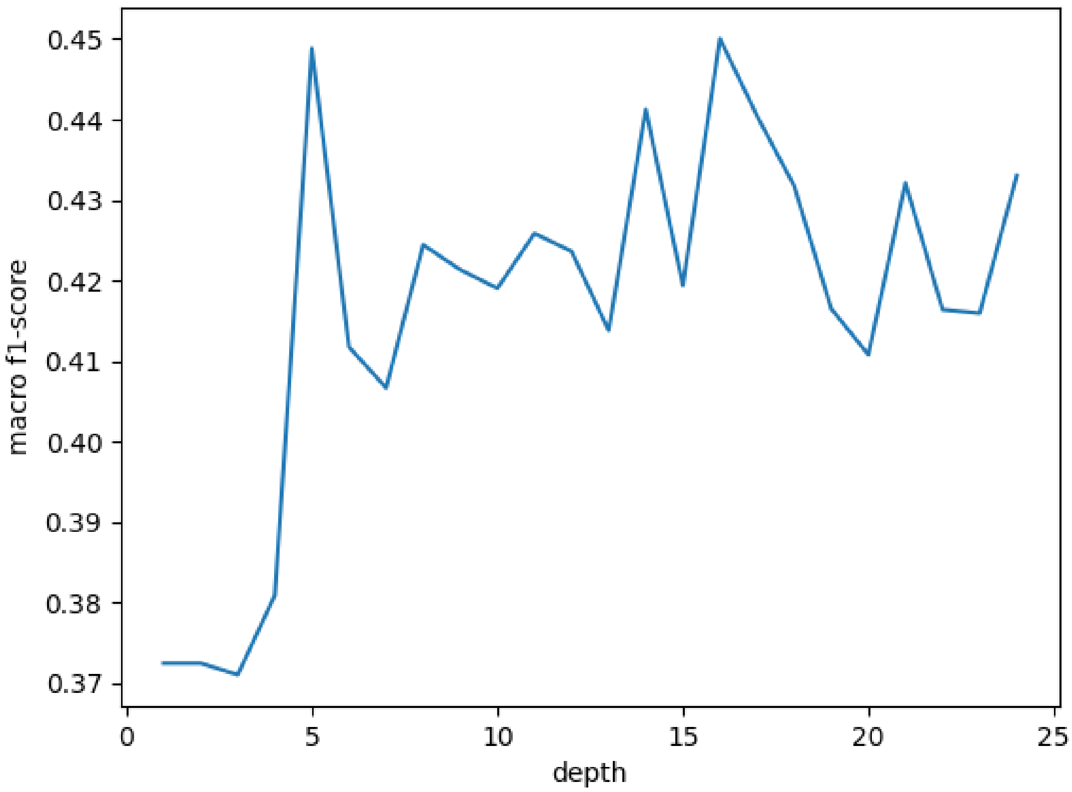
	precision	recall	f1-score	support
High	0.47	0.08	0.13	91
Low	0.88	0.98	0.92	1108
Mid	0.60	0.50	0.54	183
Peak	0.00	0.00	0.00	15
accuracy			0.84	1397
macro avg	0.49	0.39	0.40	1397
weighted avg	0.81	0.84	0.81	1397

Table 1: k-NN classification report

Decision Tree

Then the python sklearn DecisionTreeClassifier is used to build a decision tree to predict the power consumption of the next 30 minutes. For a tree the data does not have to be scaled. The dataset is again divided in 80 percent training data and 20 percent test data. An optimal depth is found by calculating the accuracy for different depths. The accuracy is the highest for depth equal to 5 and also the F1 score has a peak at that value. A higher depth however shows better results in separating the classes with lower support, while depth 5 works very in labelling ‘Low’. This can be seen in table 2, calculated with depth 5.


0.85
0.84
0.83
0.82
0.81
0.80
accuracy
0
5
10
15
20
25
depth


0.45
0.44
0.43
0.42
0.41
0.40
0.39
0.38
0.37
macro f1-score
0
5
10
15
20
25
depth

In table 1, the metrics of the classification can be found.

	precision	recall	f1-score	support
High	0.53	0.09	0.15	91
Low	0.90	0.96	0.93	1108
Mid	0.57	0.61	0.59	183
Peak	1.00	0.07	0.12	15
accuracy			0.85	1397
macro avg	0.75	0.43	0.45	1397
weighted avg	0.83	0.85	0.83	1397

Table 2: Decision Tree classification report

Conclusion

Both models perform well in labelling the most supported class ‘Low’. This can be seen in the high f1-scores of 92% and 93%. As a result of the fact that this label is the most prevalent, the general accuracy is high with 84% and 85%. The performance for less prevalent classes is not good. The k-NN algorithm doesn’t predict any peaks, because of the low support of this class and the large amount of nearest neighbors. The decision tree performs a little bit better and with a higher depth the results for the other classes can be improved while the accuracy of the ‘Low’ class decreases. The high consuming periods are however the most interesting to predict. If the goal is to predict periods of high power consumption, more data is needed. Both in amount of features, as in period of time in order to have more support in the higher consumption classes.