Comparing different approaches for Human Occupancy Detection based on eletricity consumption

Kai Ketzler
School of Business and Economics
Humboldt University of Berlin
ketzlerk@hu-berlin.de

Muhammad Najati Al-imam
School of Business and Economics
Humboldt University of Berlin
muhammad.najati.alimam@hu-berlin.de

David Steiner
School of Business and Economics
Humboldt University of Berlin
steineda@hu-berlin.de

March 25, 2021

Abstract

The focus of today's innovation and technology is shifting towards building more eco-friendly environments. There is a growing effort to reduce energy consumption and carbon footprint. Human Occupancy Detection is one of the leading topics of discussion and research when it comes to optimizing energy consumption of households or buildings.

This study introduces the topic of Human Occupancy Detection and follows a comprehensive approach for detecting human occupancy within a household in a non-intrusive way using energy consumption in Watts. In order to achieve this, a variety of models and features derived from the Watt consumption have been implemented and tested on serveral datasets. To get the best performing model possible for this task, different combinations of trained models and aggregation times have been compared and evaluated by the predictive accuracy. The aim of this research is to provide a model that is robust and can reliably predict human occupancy in a variety of households with different energy consumption patterns.

***K*eywords** Occupancy, Human Occupancy Detection, Occupancy Prediction, Electricity Consumption, Smart Homes

1 Introduction

In the past few decades, there has been extensive research and advancement in the field of occupancy detection. That is, detecting if a household, an office or a building is occupied. The interest in this topic stems from its importance as a baseline for building more energy efficient smart homes and buildings. For example, Human Occupancy Detection can be used in developing smart heating and ventilation systems that regulate temperature, smart lighting systems or to automate some appliances based on occupancy. The increasing number of appliances in households shows how relevant this topic is and even more so in the future. Especially considering the urgent need to mitigate the effects of global warming. There are multiple ways of detecting human occupation. Some of which are highly intrusive, like cameras. Others are not as intrusive, but can be expensive to implement, like installing sensors inside a household. One of the low or non-intrusive alternatives that is easy to implement while being cost-efficient is occupancy detection based on total energy (Watt) consumption, which is the method of detection used in this research.

We present the state of research to date and give insight on how this paper differs from the previous work in this field. Afterwards, we describe the scientific approach and methodology with which we want to achieve the goal of creating a model that is able to reliably predict human occupancy in households using only the Watt readings. The used datasets and how they were transformed in order to harmonize them are described together with the necessary data cleaning and the applied feature engineering. With the foundation laid, building of the models and evaluation of the results achieved by the different models are discussed. We conclude the paper by discussing the progress overall and suggested improvements for future research.

2 Related Work

In this chapter we will present the current state of the art research in the field of human occupancy detection as well as most related approaches to our work.

Occupancy Detection is a broad field and can be categorized into different subgroups as shown in figure 1:

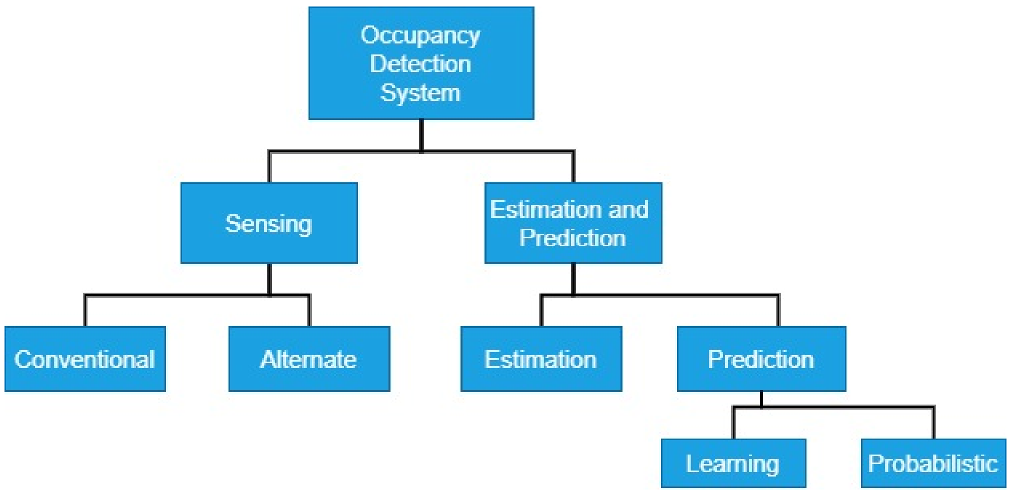

Figure 1: Overview Human Occupancy Detection (Trivedi and Badarla, 2019).

Trivedi and Badarla [5] provide with their work a survey of occupancy detection systems, as well as an overview of the available occupancy detection approaches while highlighting the underlying limitations and issues.

Sensing deals with, as the name suggests, sensing in real time whether a building or a room are occupied and can be split into conventional sensing (using sensors e.g. infrared sensor), and alternate sensing (using smart devices, network activity or electrical energy usage). The Estimation and Prediction group focuses on future events based on past data and tracks an occupants behaviour in order to predict or estimate the occupation status. The prediction subgroup can be divided into prediction using learning, and probabilistic prediction. Our work focuses on both as we want to achieve the best prediction possible for the given problem.

Occupancy detection systems can use different methods in order predict possible occupation. Some occupation estimation systems use equipment such as mobile phones or radio frequency identification tags which are carried by end users. Other systems make use of sensors such as acoustic, passive infra-red (PIR), motion sensors and depth cameras, which are capable of producing occupancy information without the usage of devices. However, occupancy prediction is also possible using measured power consumption. This paper deals with occupancy detection of residential homes using SMART meters which record the electricity consumption in small time frames. Kleiminger et al.[2] derive occupancy information from electric load curves measured by smart meters using the ECO dataset with supervised machine learning algorithms. Using 35 features, principal component analysis (PCA) and feature selection together with a set of 7 classifiers, Kleiminger et al. managed to achieve a classification accuracy of up to 94%. Their work also shows, that no single feature set performs consistently well over all households. The applied machine learning algorithms were support vector machines (SVM), K-nearest neighbours(KNN), Gaussian mixture models (GMM), hidden Markov models (HMM), and a simple threshold approach. Regarding the limited amount of used algorithms in their research, we want to extend the search for an optimal learning algorithm and use only the best performing ones from previous research in combination with a set of other (untested) machine learning algorithms.

Extending their previous research, Kleiminger and Becker[1] explored the possibilities of zero-training algorithms for occupancy detection on multiple datasets. They suggest that smart meters become more widespread in households, thus making the adaption of human detection algorithms in the everyday life easier, but as the data is unlabelled, the algorithms might not be suited for real world settings. For this reason, the Hidden Markov Model, Gemoetric Moving Average, the Page Hinkley test, and Non-intrusive Occupancy Monitoring (NIOM) were tested on three publicly available datasets and compared to supervised algorithms (Random Forest (RF), KNN and SVM). While sampling on 30-minute intervals, the best performing algorithms showed an accuracy of 69–90%. Extending this research even further, we aim to combine the approaches of Kleiminger et al.[2] and Kleiminger and Becker[1] in order to find a suitable supervised algorithm for multiple datasets.

Vafeiadis et al. [3] investigated the problem of occupancy detection, as well by using machine learning algorithms in combination with Monte Carlo simulations on energy and water consumption data of a household in order to find the classification technique with the highest F-measure. The algorithms analyzed were SVM, Decision Trees (DT), RF and Back-Propagation Network (BPN) along with the AdaBoost algorithm. Mutual Information was used for feature selection in an effort to reduce the dataset's sparsity and to retain classification performance. Random Forest and Boosting performed best with an F-measure of 83.37% and 82.79%, respectively. With the gained knowledge of the performance for supervised algorithms, we take RF and Boosting into the analysis as well as state of the art machine learning algorithms that have not been analyzed yet while considering solely electricity consumption data.

While the set task is a combination of multiple approaches from different works, we see our work as a generalization and expansion of multiple works in this area.

3 Methodology

To achieve the goal of finding a robust human occupancy detection algorithm, we conducted a literature review and pursued a systematic mapping procedure. This contains finding suitable papers and classifying them accordingly to the defined research question. As our research question deals with the subject matter which model predicts human occupancy best and generalizes well over multiple datasets using only the electric power usage of a household, we narrowed the search down to different methods of human occupancy detection as well as overall approaches in this area of research.

In order to gather the papers, we identified relevant search terms and searched on major academic platforms for suitable results (Google Scholar, ResearchGate, JSTOR, Springer, ScienceDirect, IEEE). We classified the papers by their applied algorithms, achieved prediction accuracy and uniqueness of approach with the aim to encompass the broad field of research and be able to identify the current state of the art. Therefore we analysed around 20 papers and reviewed the approaches and findings. After the screening process we decided to lay the focus on a smaller amount of papers which were discussed in the chapter Related Work. While conducting the literature review further questions for our set task arose which were solved differently in various approaches:

- Which models do we chose to compare?
- Will the model learn while handling data or is it based on past data?
- Which time window is the best to aggregate on?
- Which measure do we use to evaluate the model?
- Could external factors such as the outside temperature improve the predictive capabilities?

As all these questions were answered differently in the screened works, we had no choice but to resolve them by ourselves based on our gained knowledge. We discuss this questions in their respective chapter and disclose our choices there.

After analysing related work we started to search for datasets that were suitable for the task. As we analysed multiple sources we found three datasets that provided promising data for our goal. In order to achieve our goal of creating a robust model that can reliably predict human occupancy we set out the following approach get the best result:

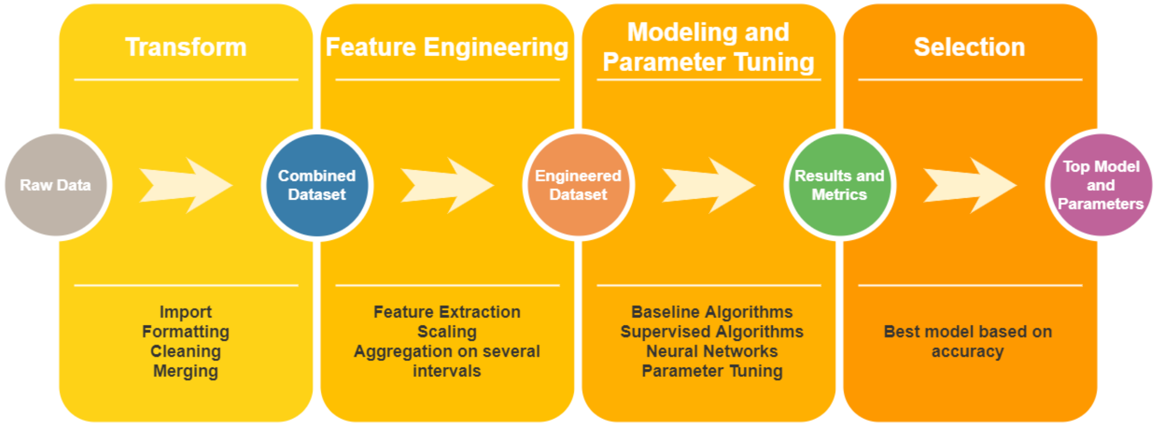

Figure 2: Data to model process.

4 Description of the Datasets

In order to accomplish the goal of providing a robust model that can cope with a variety of watt usage patterns, different datasets where needed as the datasets found contained only a few households. With a combined dataset not only the amount of households increases but also the amount of recorded electricity usage and therefore possible patterns that can be learned by models. For this reason this paper utilizes multiple different publicly available datasets which are described in the chapters below.

As the research showed, suitable datasets in the area of human occupancy are very rare, because most datasets do not have the required ground truth, that states if a household was occupied or not. Therefore, this paper will focus on three datasets that where suited for the described problem.

4.1 DRED

The publicly available Dutch Residential Energy Dataset (DRED) contains occupancy and electricity consumption data for one household located in the Netherlands. The data was recorded during Winter and Summer with a combined duration of 6 months. The electricity usage was recorded every second using smart meters, whereas the occupation was measured using motion sensors. Whenever a person in the household moved from one room into another, a motion sensor fired a signal and recorded the date and the time. This can lead to errors or the misunderstanding of the data as the person in the household could occupy the house but without movement the gap between firings would suggest an unoccupied house. For example, if the person in the household is sleeping, no movement is recorded and thus the data would suggest an unoccupied home.

As this type of recording can not be trusted for true ground truth data, the data needs to be cleaned and extensively worked on in order to be used as the basis for the target variable prediction. The conducted transformation will be explained in chapter 4.4.

4.2 Smart*

Smart* provides power data for 3 households located in the US in Western Massachusetts using plug meters. The electricity data was measured for every appliance in the household and was recorded over a period of 3 weeks. The different electricity readings where aggregated on one minute although aggregated data is only available for two households in the seasons Spring and Summer. The data for the ground truth was recorded by using the GPS traces of the inhabitant's smart phones. Due to the 1 minute aggregated data, the DRED and ECO data have to be aggregated to 1 minute as well in order to harmonize the datasets.

4.3 ECO

The Electricity Consumption and Occupancy (ECO) dataset is publicly available and the largest one considering the household size and recording time. It contains both ground truth and electricity consumption data for six households over a duration of 8 months. The participating households where located in Switzerland and recorded their occupation using a tablet every time they left or arrived at their home. The ECO dataset provides 1 Hz aggregate consumption data (current, voltage, and phase shift for each of the three phases in the household) and also 1 Hz plug-level data measured from selected appliances.

Additionally it is important to mention that the ground truth data was only available for five households, therefore we only used these for the final dataset. Also the ground truth occupation data has gaps in the recording because the occupants in some cases forgot to track their occupation status. The authors of the paper corresponding to the dataset zeroed the complete day if a missing tracking was recognized.

Looking at this, we can see that the capture frequency is not always the same. The DRED and ECO datasets have Meter Frequency readings every second. That means we will have to aggregate them to the same base level as the Smart* dataset (1 minute)

4.4 Data Transformation

In order to create a combined dataset, the present data needs to be harmonized. That means bringing together data of varying file formats, naming conventions, and columns, and transforming it into one cohesive dataset. The following graphic summarizes the three datasets that were combined.

Dataset	No. Households	Country	Capture Frequency	Duration of recording	Seasons
DRED	1	Netherlands	1 Second	6 Months	Summer/Winter
Smart*	2	U.S	1 Minute	3 Weeks	House1: Spring/Summer House2: Summer
ECO	5	Switzerland	1 Second	8 Months	Summer/Winter

Table 1: Overview of used datasets.

After merging the datasets into one, data cleaning was performed. This is an equally important step since the performance of the model highly depends on the quality of data provided.

Some data cleaning was done beforehand during the import of the data. There are some minor differences in the way the datetime variable is represented, which was a simple step to unify across datasets early on. The 3 datasets are generally reliable when it comes to formatting and datatypes as the data at hand is relevant, correctly represented, and no duplicates exist. However, NA removal, outlier truncation and scaling were still necessary to achieve good results.

The datasets contains substantial outliers which are caused by certain appliances. For example an oven or a washing machine consume a lot of power during a short period of time, while other appliances consume a lot of energy during startup. There could be several reasons for these power spikes, but they do not represent normal energy consumption and could disorientate the model. Since the main concern is only about the prediction of the occupancy status, outliers were truncated based on the Z-score instead of being completely removed.

5 Feature Engineering

Feature engineering is an essential step in building a good model. In the case of non-intrusive Human Occupancy Detection, we have to rely heavily on very few inputs, mainly the energy consumption. However, the date-time is very useful as well.

With the electric consumption data an extensive set of features can be extracted. Therefore, findings from related work in the field were analysed, which helped to identify important features for model prediction. In addition we extracted further features that could be useful. The relevance of some of these features is hard to evaluate beforehand, however, many different features do not harm the model as most models perform feature selection. Additionally feature importance can be derived from the models afterwards.

In addition to features derived from the electricity consumption, we want to experiment with sunlight as a feature. Being able to tell whether it is daytime outside can help the model identify when to expect the increased energy consumption originating from illumination devices. This can be achieved with the help of an API that provides sunrise and sunset times given a geographical location and a date. Based on which, a new feature is engineered to indicate whether it is day-time (1) or night-time (0). It is worth mentioning that the geographical location provided to the API is based on the country. The datasets used in this research do not specify in which cities the households are located, therefore, this feature is not optimal in its current state.

The following list summarizes the engineered features with a short description:

Feature	Description
Watt min	Minimum value during aggregation time
Watt max	Maximum value during aggregation time
Watt mean	Mean value during aggregation time
Watt std	Standard deviation during aggregation time
Watt last	Last watt reading in previous aggregation time
Watt mean percentage change	Change in watt mean percentage between last instance and current one
Watt mean absolute difference	Absolute difference in watt mean between last instance and current one
Watt sum of absolute differences	Sum of absolute differences up to current instance
Watt rolling 3	Rolling average of watt over last 3 instances
Watt rolling 5	Rolling average of watt over last 5 instances
Watt rolling 10	Rolling average of watt over last 1 instances
Watt cumulative sum 3	Cumulative sum of watt over last 3 instances
Watt cumulative sum 5	Cumulative sum of watt over last 5 instances
Watt cumulative sum 10	Cumulative sum of watt over last 1 instances0
Watt log	Base-2 logarithm of Watt
Sunday to Saturday	Dummy variable for each day
Is weekend	Dummy. (1) if current day is Saturday or Sunday. (0) otherwise
Sleeptime	Dummy. (1) if current time is between 23:00 and 06:00. (0) otherwise
Sunlight	Based on API. (1) if it is day time during this instance. (0) otherwise

Table 2: List of extracted features.

Another crucial step is scaling as it brings all features in the same standing in order to prevent one significant number impacting the model because of their large magnitude. The Min-Max scaler is our scaler of choice and yielded the best results concerning model performance. After application of the scaler, the data is rescaled such that all feature values are in the range of 0 to 1.

6 Prediction Models

This section copes with the prediction of the human occupancy status in different households of the combined dataset. We first setup a baseline after which we use supervised learning to show the potential gain. After that we compare the best supervised model to a Neural network.

Evaluation Metrics Even though the predictions are made on several time intervals, we will be evaluating the performance on the 30-minute interval. This makes the results comparable and more in-line with other work in the field. It is also a fitting time interval for a real life application as the time interval neglects short absences but captures an unoccupied home rather quickly.

For the performance evaluation we focused on the accuracy as this metric was referenced in most other papers. However, due to the the imbalance of the dataset, that contains more data points where the household was occupied, we regarded the F1 score as well. In order to compare different models AUC was used.

Accuracy is also commonly used in prediction tasks as it evaluates binary classification models by summarizing the True Positives and the True Negatives and dividing it by the sum of all the predictions made. Although it is a poor evaluation metric for heavily biased data, it is used in most related research, which is why we give it the greatest importance while evaluating.

The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. The formula for the F1 score is:

F1 = 2 * (precision * recall) / (precision + recall) In the multi-class and multi-label case, this is the average of the F1 score of each class with weighting depending on the average parameter.

Finally, the AUC score is used to rank and select the best models.

6.1 Baseline Statistics / Predictors

Before creating sophisticated models, we attempt to predict the occupancy with simple statistics using thresholds. This serves as a baseline to compare against. We decided to implement three different baselines:

- Baseline: The household is always occupied
- Threshold Baseline: A certain Watt threshold needs to be crossed
- Moving Threshold: A moving average needs to be crossed

The first baseline "predicts" that the household is always occupied as this class has the biggest share in the dataset. The second baseline introduces a threshold, which needs to be crossed, after which the household is marked as occupied. While this seems promising at first the results were quite disappointing as the threshold is hard to estimate because of the high fluctuations in the electricity consumption. For this reason we implemented a third baselines which takes past data into consideration as the third baseline represents a moving average threshold that needs to be crossed in order to classify the household as occupied. The following figure shows an excerpt of the data, where clear spikes in the Watt usage are detectable.

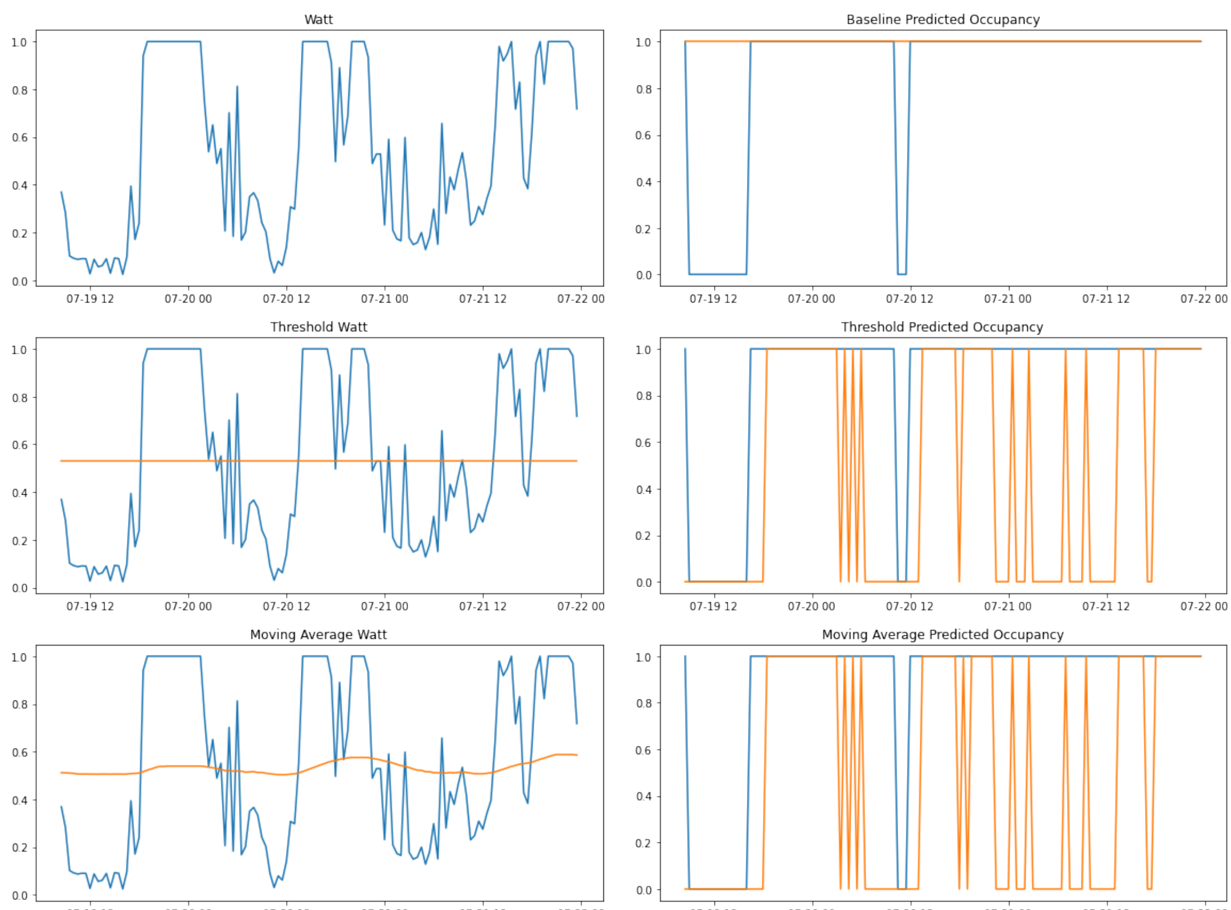

Figure 3: Baseline moving average on a 30-minute interval.

The blue line represents the measured electricity consumption of one household during one week while the orange line displays the created threshold which is used for prediction. The three plots on the right side show the ground truth data in blue and the predicted occupancy in orange. The predicted results are far from perfect. Interestingly it can be seen that the predicted occupancy fluctuates strongly for the second and third baseline. This is caused by a low electricity measurement while the occupants are at home. During the day it seem to occur quite often that the electricity consumption drops below the threshold which can be caused by the occupant not using any electrical device connected to the current output. Out of the three, the first baseline performed best with an accuracy of 81% on a 30-minute time interval.

The following table shows the performance metrics of the baselines for the whole dataset with seven different aggregation times:

Simple Predictors			
Model	Accuracy	F1	Agregation Time (mins)
Baseline	0.796568	0.886767	5
Threshold	0.589731	0.683572	5
Moving Average	0.499473	0.596913	5
Baseline	0.799606	0.888646	10
Threshold	0.589514	0.684146	10
Moving Average	0.549949	0.656498	10
Baseline	0.802552	0.890462	15
Threshold	0.594278	0.688518	15
Moving Average	0.503259	0.604995	15
Baseline	0.811353	0.895853	30
Threshold	0.60711	0.700389	30
Moving Average	0.512116	0.621196	30
Baseline	0.819578	0.900844	45
Threshold	0.617308	0.709989	45
Moving Average	0.511516	0.623504	45
Baseline	0.828132	0.905987	60
Threshold	0.619501	0.713508	60
Moving Average	0.519029	0.631781	60
Baseline	0.845013	0.915997	90
Threshold	0.617146	0.715353	90
Moving Average	0.533403	0.657052	90

Table 3: Baseline performance metrics.

As seen above, the best performing method is the baseline prediction where the household is always set to occupied. It is important to keep in mind that accuracy improves for bigger time aggregations. If the household was occupied at least once within the time interval the whole aggregated time interval will be set to occupied. For example, if a person was at home and went grocery shopping for half an hour in the two hours aggregated, it would seem like they never left the house.

6.2 Training and Testing Split

Related work often used only one household and evaluated the models with cross validation. This approach is suitable if the goal is to predict further occupation based on previously gathered data for a specific household. However, it might be hard to apply in real world settings, since it would require to track a households occupancy status beforehand. Since our goal is to build a classifier that can be deployed instantly without requiring further training data from specific households, it is important to include several households with diverse consumption patterns in the training data that the model can learn from. The extensive EDA of our data showed that this is the case for the collected data.

In our approach we split the data into a training and evaluation set with the majority of the households assigned to the training set and the remaining in the evaluation set. Since the biggest diversity in the consumption pattern was among the different datasets (DRED, SMART*, ECO) it was important to include at least one household from each dataset in the training data. Owning to the fact that we only have one household in the DRED dataset, we assigned the first 80 percent of the tracked days to the training data and the remaining to the test data. For further reference we called this train/test split "combined".

Furthermore we wanted to evaluate the predictive capabilities among the different datasets (e.g using the DRED data to predict the household from the SMART* data). Accordingly we created further training/test sets that include all households from only one dataset. We gave these trainig and test splits the name of the equivalent dataset (Dred, ECO, Smart*).

6.3 Supervised Machine Learning

Continuing the search for the optimal model considering human occupancy detection, supervised models can bring a significant gain as supervised learning is a powerful tool to classify and process data using machine learning. The learning of the models is achieved by using some data for prediction and different data for the classification of unlabeled data through which the model tries to minimize the predicted error.

For the prediction task we used a variety of models. By evaluating the performance of multiple models using the accuracy, f1-score and AUC we were able to identify the best performing ones. The following list describes the tested models:

- Logistic Regression
- Bagging
- AdaBoosting
- RandomForest
- GradientBoosting (sklearn)
- LightGBM (Microsoft)
- XGBoost

The following figure displays the performance of the supervised models on all time intervals when training and predicting on the combined dataset:

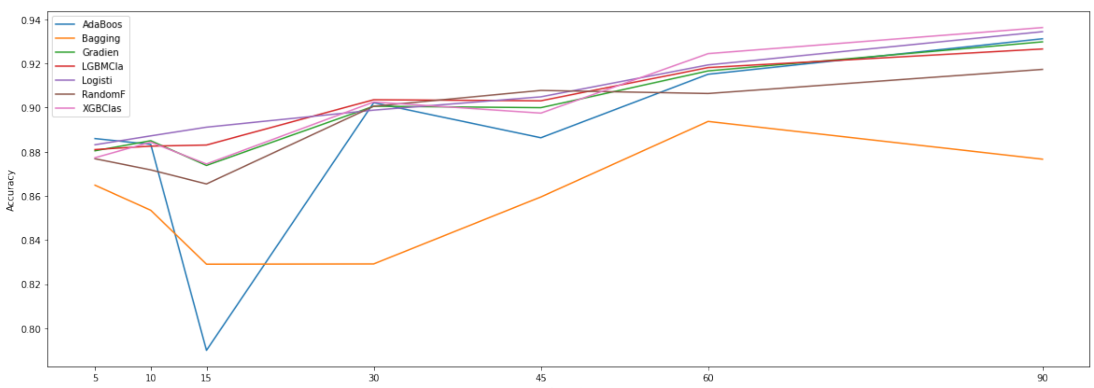

Figure 4: Supervised model accuracy per time interval.

While XGBoost performs better on the 60-minute interval and upwards, LightGBM has higher accuracy on the 30-minute interval. That, in addition to being relatively fast to train, makes it the proposed optimal model for the 30-minute timeframe.

To put this in numbers, the table below demonstrates the performance of the models on the 30-minute interval. Keep in mind that this is prior to hyperparameter tuning:

Model	Aggregation time	Acc train	Auc	F1	Train dataset	Test dataset	Acc test	Auc test	F1 test
Logistic Regression	30	0.785732	0.745689	0.879877	Comb	Comb	0.898822	0.695012	0.946674
Bagging	30	0.915735	0.948095	0.946454	Comb	Comb	0.833726	0.563712	0.907193
AdaBoost	30	0.869998	0.901437	0.92014	Comb	Comb	0.902209	0.725135	0.948415
RandomForest	30	0.934626	0.974867	0.959079	Comb	Comb	0.899853	0.613905	0.947082
GradientBoosting	30	0.896595	0.938187	0.936118	Comb	Comb	0.900589	0.619032	0.947483
LGBM	30	0.924435	0.967527	0.952441	Comb	Comb	0.903535	0.680404	0.948968
XGB	30	0.895352	0.938446	0.935261	Comb	Comb	0.902504	0.653644	0.948571

Table 4: Supervised models performance metrics.

Although we are interested in the combined dataset, it would be interesting to see how each of the 3 datasets perform when being trained to predict on the combined one. The plots and performance table for these can be seen in *Appendix A* and *Appendix B* respectively.

6.4 Model performance and optimization

After settling for the overall best performing model, hyperparameter tuning is necessary to further increase the accuracy. In order to find the optimal parameters a GridSearch was performed after which the parameters were updated. For our application we used following parameters: colsample_bytree: 0.9, max_depth: 6, n_estimators: 110, num_leaves: 100. With the updated parameters the model achieved better results than before as can be seen in the following table:

	Percision	Recall	F1 Score	Support
0	0.87	0.78	0.82	864
1	0.94	0.97	0.95	3159

Table 5: Performance metrics for tuned LGBM.

After the hyperparameter optimization, the model was able to predict with an accuracy of over 93%. The individual model performance for different households of the test set can be seen in the following figure.

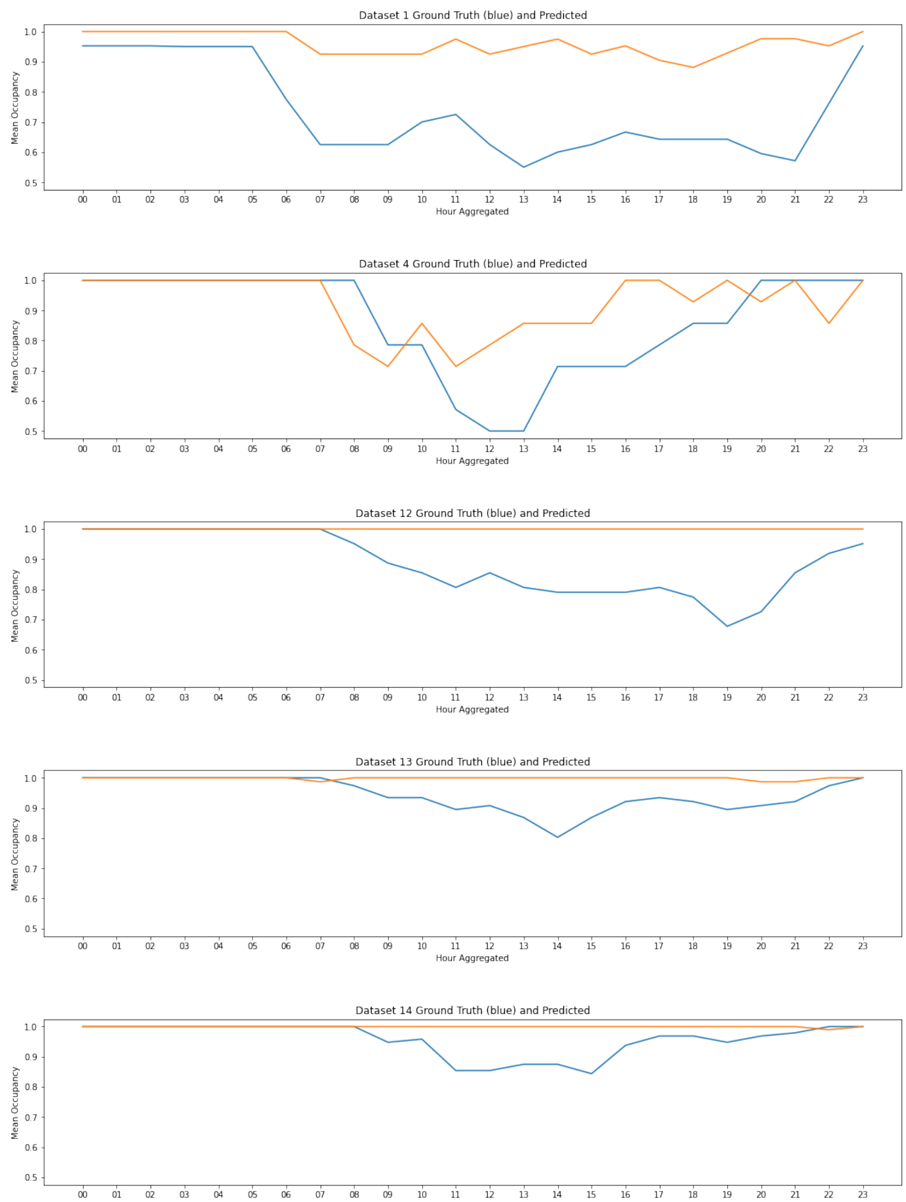

Figure 5: LightGBM mean predicted occupancy per hour.

The figure shows the average occupancy (blue) during the different hours of the day. All available days for the household are aggregated and the mean occupancy displayed. The orange line shows the predicted occupancy during the time of the day. As it can be seen the model predicts a higher occupancy rate than the ground truth data and even starts to resemble the baseline in cases with a high average occupancy rate. In other cases the model captures the occupancy status rather well. Overall this gives an idea of how the model predicts although it is worth mentioning that individual days need to be analysed as well in order to fully understand model performance. For example, the model could capture two days of absence in the week which would not occur in the figure as the figure displays the mean occupancy over the whole data available for the household. The model seems to prefer predicting occupied which in real life scenario comes at smaller misclassification cost than predicting unoccupied (assuming for example, that turning off the heating, while the house is occupied, is worse than turning it on while the house is not occupied).

Feature Importance In general, each of the supervised models assigned different importance for features. However the top three features stayed mostly the same. The selected model identified following feature importance :

Feature	Importance
Watt Min	633
Watt rolling 10	570
Watt sum of absolute differences	518
Watt rolling 3	306
Watt Std	304
Watt Max	285
Watt rolling 5	269
Watt cumulative sum 10	258
Watt mean absolute difference	242
Watt Last	208

Table 6: Most important features for LGBM.

6.5 Neural Network

To see if we can further increase the predictive capability, we created a feed forward neural network. The best performing structure that we could identify comprises of 4 Dense layers, 3 relu and one (in the end, due to the binary classification problem) sigmoid activation function. We decided to implement a Keras feed forward neural (sequential) network, since we want to recognize patterns in the data, which makes this network structure suitable for this prediction task. Our optimizer of choice is Adam with a learning rate at the default Adam setting of 0,001 [9]. To avoid overfitting, a Dropout rate is implemented, varying in-between the first three dense layers from 10 to 20 percent. Considering the batch size, lower sizes are usually preferable to higher ones [10], which is why we chose a low setting of 5.

After training for 80 epochs, the proposed neural network achieved an accuracy on the validation test set of 89%. While the model accuracy converges after epoch 10 the validation loss seems to reach a local minimum at epoch 35 and starts overfitting afterwards as seen in the figure below.

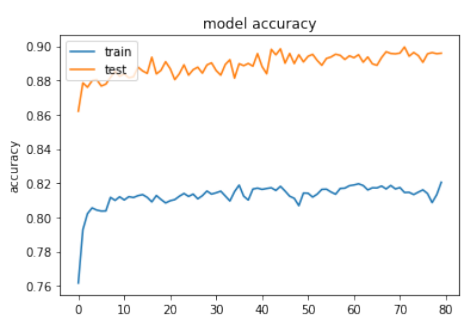

Figure 6: Neural Network accuracy.

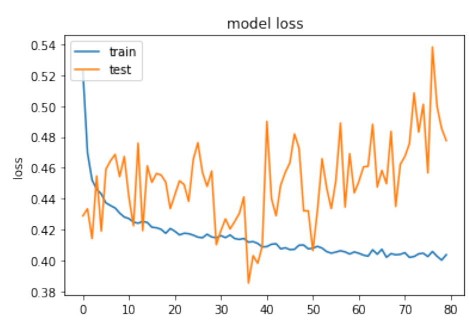

Figure 7: Neural Network loss.

Regarding predictions on individual households in the datasets *(Appendix C)* the Neural Network captures households with a varying mean occupancy rate very well but struggles in settings with households that have a consistently high occupancy rate (above 60%). In some cases it even struggles to predict an occupied household during night hours, in which the occupancy status is evident in the data. The neural network managed to achieve an AUC of 72% and a validation loss of 0.48 meaning it predicts rather well but has a high uncertainty about its predictions. This can be due to miscalibration or the uneven class distribution. A possible solution for this would be to oversample the underrepresented class and review the architecture and parameter settings of the model.

7 Further Research

Expanding our research in the future, we suggest using more diverse data as well as different datasets. This could prove if the provided model is able to reliably predict the occupancy status. Using more data that contains a lower occupation rate could also help the model to perform better overall. In addition a higher amount of publicly available datasets with ground truth occupancy data would tremendously help to further increase the research in this area. Also transforming and combining the different datasets resulted in a high effort which is why a standardised format for electric power consumption would be suggested. A standardized dataset would further help significantly to compare the results of related work, since often different data transformation and cleaning methods were used.

In order to boost performance several ideas could be explored. For example, including more external features like outside temperature, which can be achieved by using an API. The usage of Non-Intrusive Load Dis-aggregation Algorithms can also help with the prediction accuracy as they could predict which device is currently put to use. This information could then help the model predict the occupation status, as it shows if a certain device is turned on, that wouldn't be turned on without human interaction (e.g. lights, TV). This however requires in-depth research because many edge cases need to be avoided (e.g. turning on the TV to record a show, or setting a timer on the washing machine). In addition, more external features like the exact geographical location of the households similar to our engineered feature "sun" could be implemented. The sun feature in it's current state has just the country's location in which the household is located, which is prone to fail in countries with many timezones. We also see further potential in developing more complex neural network structures with different layers or parameter settings.

8 Conclusion

The purpose of this paper was to follow a wide approach to the problem of human occupancy detection by combining several datasets from different sources. The aim was to derive the highest possible prediction-accuracy of a household's occupancy status using only minute-based aggregation of energy consumption. With the help of an extensive set of extracted features and a variety of models deployed, we were able to predict human occupancy on a combined dataset with an accuracy of **93%**. Overall boosted tree based models performed best in this task. This performance can be increased further with the inclusion of more diverse data, as many models had problems predicting the occupancy for households with a high occupancy rate.

We first explored the different datasets and transformed the data in order to combine and aggregate the tracked energy consumption time into the equally smallest time possible (1min). From there we formed different aggregation times which were later used to predict the occupancy status. With this done, the unified dataset was cleaned before new features were engineered solely out of the electricity readings. After that, we set a baseline prediction for the occupancy status with zero training models and compared them to a variety of supervised models and a neural network. Out of the used algorithms we were able to show that supervised prediction models outperform all zero training algorithms and even outperformed a feed forward Neural Network in this task.

Considering the achieved accuracy, the best baseline was already able to achieve an accuracy of 81% whereas the best supervised model (LightGBM) was able to achieve 93% after hyperparameter tuning. The feed forward neural network only managed to get an accuracy of 89%. Here we can also see the downsight of using only a singular evaluation method as the accuracy alone is highly dependent of the data. Regarding the overall prediction of human occupancy, metrics like the F1 score, recall and precision play an important role. Plotting the aggegrated mean predictions similar to *Figure 7* or *Appendix C* can also give a good understanding of the overall model performance.

It is worth mentioning, that the trained models assigned different importance levels for the features given, however, a few of these features reoccur among the top predictors across different models. All of the most important features were derived from the electric power consumption. The best identified features for human occupancy detection were: The minimum Watt for the given time, The average Watt of the last 10 readings and the sum of absolute differences between previous instances up to the current one.

In comparison to related work, we managed to achieve an improvement of the prediction accuracy compared to the work of Kleiminger et al(2017) on a combined dataset, however comparing to the work of Kleiminger et al. (2015), our model achieved a worse performance (Kleiminger 94%, our model 93%). However we trained the model on three datasets combined while the work of Kleiminger et al. (2015) only uses the ECO data. Also the accuracy of Kleiminger et al(2015) refers to a single household during wintertime while the same model predicted much worse on other households. Since our model validated on unknown data and multiple households we assume that the loss of accuracy is negligible and could even be better suited for a real-life scenario.

Additionally we want to point out that Human Occupancy detection introduces several advantages for smart homes and buildings, such as reducing energy consumption and increasing indoors comfort. On the other side, deploying occupancy based algorithms can also be a major risk and potentially be exploited by malicious entities.

References

[1] V. Becker, W. Kleiminger: *Exploring zero-training algorithms for occupancy detection based on smart meter measurements*, 2017.

[2] Wilhelm Kleiminger, Christian Beckel, Silvia Santini: *Household Occupancy Monitoring Using Electricity Meters*, 2015.

[3] T. Vafeiadis, S. Zikos, G. Stavropoulos, D. Ioannidis, S. Krinidis, D. Tzovaras, K. Moustakas: *Machine Learning Based Occupancy Detection Via The Use of Smart Meters*, 2017.

[4] Verma, Singh, Majumdar: *Multi Label Restricted Boltzmann Machine for Non-intrusive Load Monitoring*, 2019.

[5] Dipti Trivedi, Venkataramana Badarla: *Occupancy detection systems for indoor environments: A survey of approaches and methods*, 2019.

[6] Davide Pereira: *Occupancy Prediction from Electricity Consumption Data in Smart Homes*, 2017.

[7] Klemenjak et al.: *On performance evaluation and machine learning approaches in non-intrusive load monitoring*, 2018.

[8] Sunrise-Sunset, *sunset and sunrise API*, retrieved from: https://sunrise-sunset.org/api.

[9] Kingma, D. P.; Ba, J. *Adam: A method for stochastic optimization*, 2017.

[10] Keskar, N. S.; Mudigere, D.; Nocedal, J.; Smelyanskiy, M.; Tang, P. T. P. *On large-batch training for deep learning: Generalization gap and sharp minima*, 2017.

[11] Delft University of Technology, *Dutch Residential Energy Dataset (DRED)*, retrieved from: https://www.tudelft.nl/en/eemcs/the-faculty/departments/software-technology/ akshay/dred.

[12] Eidgenössische Technische Hochschule Zürich, *Electricity Consumption & Occupancy (ECO)*, retrieved from: https://www.vs.inf.ethz.ch/res/show.html?what=eco-data.

[13] Laboratory for Advanced Software Systems (LASS), *Smart* Dataset for Sustainability (Smart*)*, retrieved from: http://traces.cs.umass.edu/index.php/Smart/Smart.

[14] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, S. Santini, *The ECO dataset and the performance of non-intrusive load monitoring algorithms*, in: 1st ACM International Conference on Embedded Systems for Energy-Efficient Buildings, ACM, 2014, pp. 80-89.

[15] Z. Chen, C. Jiang, L. Xie, *Building occupancy estimation and detection: A review* , 2018.

[16] I. Mutis, A. Ambekar, V. Joshi, *Real-time space occupancy sensing and human motion analysis using deep learning for indoor air quality control*, in: Automation in Construction, Volume 116, 2020.

[17] L. Rueda, K. Agbossou, A. Cardenas, N. Henao, S. Kelouwani, *A comprehensive review of approaches to building occupancy detection*, in: Building and Environment, Volume 180, 2020.

[18] N.S. Fayed, M. Abu-Elkheir, E.M. El-Daydamony, A. Atwan, *Sensor-based occupancy detection using neutrosophic features fusion*, in: Heliyon, Volume 5, Issue 9, 2019.

[19] J. Wang, N. Chung Fai Tse, J. Yau Chung Chan, *Wi-Fi based occupancy detection in a complex indoor space under discontinuous wireless communication: A robust filtering based on event-triggered updating*, in: Building and Environment, Volume 151, 2019, pp. 228-239.

[20] L. M. Candanedo, V. Feldheim, D. Deramaix, *A methodology based on Hidden Markov Models for occupancy detection and a case study in a low energy residential building*, in: Energy and Buildings, Volume 148, 2018, pp. 327-341.

Appendix A

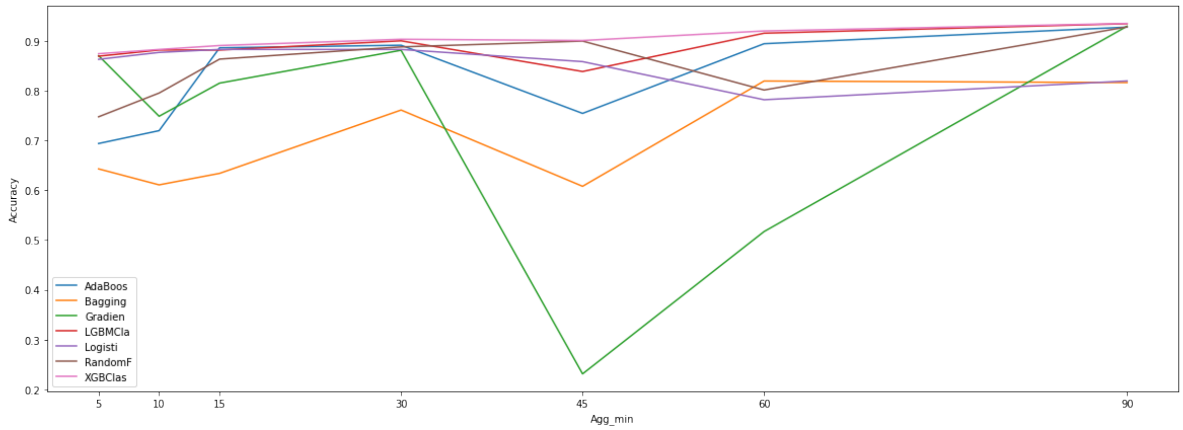

Figure 8: Supervised models - accuracy per aggregation Interval (in minutes) - Dred Training Set to Predict Combined Dataset.

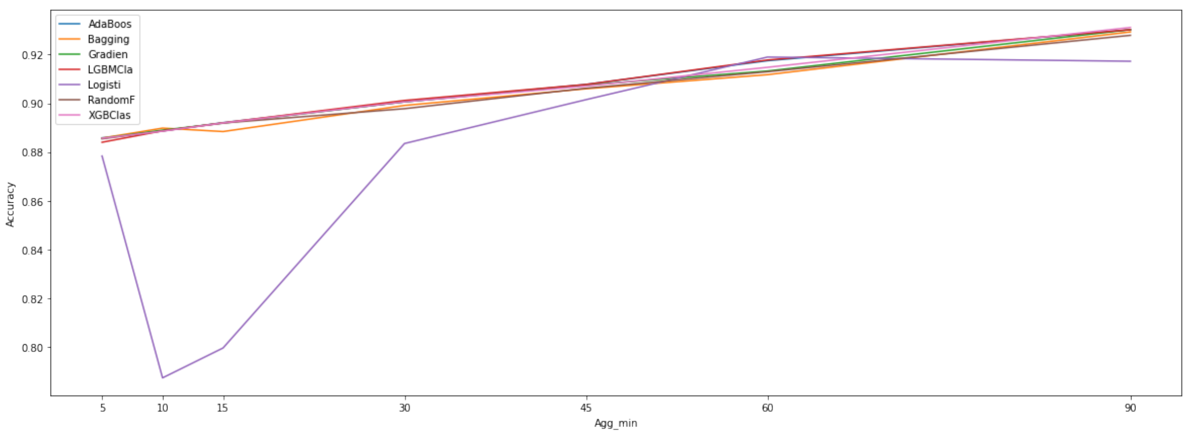

Figure 9: Supervised models - accuracy per aggregation interval (in minutes) - Smart* training set to predict combined dataset.

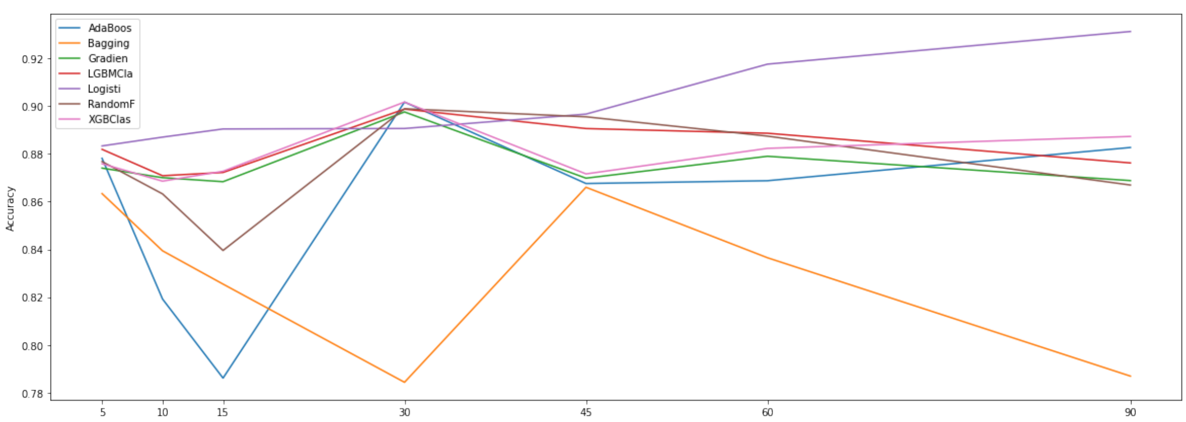

Figure 10: Supervised models - accuracy per aggregation interval (in minutes) - ECO training set to predict combined dataset.

Appendix B

Model	Aggregation time	Acc train	Auc	F1	Train dataset	Test dataset	Acc test	Auc test	F1 test
Logistic Regression	30	0.80916	0.752762	0.880763	Dred	Comb	0.884831	0.700787	0.937659
Bagging	30	0.910941	0.934225	0.940171	Dred	Comb	0.873785	0.644849	0.931037
AdaBoost	30	0.854962	0.906803	0.90404	Dred	Comb	0.891311	0.762847	0.941317
RandomForest	30	0.905852	0.963743	0.938023	Dred	Comb	0.885862	0.647993	0.938281
GradientBoosting	30	0.89313	0.944978	0.930233	Dred	Comb	0.888807	0.688039	0.940863
LGBM	30	0.894402	0.951754	0.930193	Dred	Comb	0.900442	0.719254	0.946646
XGB	30	0.885496	0.941619	0.92562	Dred	Comb	0.903535	0.739191	0.948607
Logistic Regression	30	0.763359	0.766082	0.857143	Smart*	Comb	0.883505	0.697255	0.93814
Bagging	30	0.89313	0.941082	0.928571	Smart*	Comb	0.898822	0.588571	0.946682
AdaBoost	30	0.900763	0.956871	0.933333	Smart*	Comb	0.900589	0.738675	0.947695
RandomForest	30	0.908397	0.962719	0.938776	Smart*	Comb	0.897644	0.621796	0.946061
GradientBoosting	30	0.908397	0.952632	0.938776	Smart*	Comb	0.900589	0.732881	0.947695
LGBM	30	0.931298	0.946491	0.95288	Smart*	Comb	0.901178	0.66759	0.947989
XGB	30	0.908397	0.959064	0.938776	Smart*	Comb	0.900589	0.710844	0.947695
Logistic Regression	30	0.795301	0.770166	0.885735	ECO	Comb	0.890574	0.68171	0.941894
Bagging	30	0.920502	0.95841	0.949807	ECO	Comb	0.846834	0.534739	0.915912
AdaBoost	30	0.877696	0.915449	0.924483	ECO	Comb	0.90162	0.706562	0.947918
RandomForest	30	0.933376	0.974686	0.958525	ECO	Comb	0.89838	0.579477	0.94642
GradientBoosting	30	0.905375	0.949063	0.940845	ECO	Comb	0.897496	0.635712	0.94587
LGBM	30	0.936273	0.97501	0.960016	ECO	Comb	0.898675	0.67261	0.946517
XGB	30	0.903444	0.95096	0.939467	ECO	Comb	0.90162	0.663869	0.948104
Logistic Regression	30	0.785732	0.745689	0.879877	Comb	Comb	0.898822	0.695012	0.946674
Bagging	30	0.915735	0.948095	0.946454	Comb	Comb	0.833726	0.563712	0.907193
AdaBoost	30	0.869998	0.901437	0.92014	Comb	Comb	0.902209	0.725135	0.948415
RandomForest	30	0.934626	0.974867	0.959079	Comb	Comb	0.899853	0.613905	0.947082
GradientBoosting	30	0.896595	0.938187	0.936118	Comb	Comb	0.900589	0.619032	0.947483
LGBM	30	0.924435	0.967527	0.952441	Comb	Comb	0.903535	0.680404	0.948968
XGB	30	0.895352	0.938446	0.935261	Comb	Comb	0.902504	0.653644	0.948571

Table 7: Supervised models - accuracy per 30 minutes - each dataset predicting combined.

Appendix C

Figure 11: NN accuracy per epoch.