Adaptive Function Grouping in Smart Homes

Final Report

Kazmer Nagy-Betegh

Supervisor: Dr. Cyril Renaud

Second Assessor: Dr. Anibal Fernandez

March 2019

DECLARATION

I have read and understood the College and Department's statements and guidelines concerning plagiarism.

I declare that all material described in this report is all my own work except where explicitly and individually indicated in the text. This includes ideas described in the text, figures and computer programs.

Name: ……Kázmér Nagy-Betegh……

Signature: ……Nagy-Betegh……

Date: ……21.03.2019……

Table of Contents

Table of Figures *4*

List of Tables *4*

The Abstract *4*

The Introduction *5*

Section 1: Research and Preparation *6*

Literature review 6

Hardware 6

User Interface and Software 7

Control interface comparison 7

Additional research 9

Target Audience of solution 9

Section 2: Goals and Objectives *10*

The description of the proposed idea 10

Goals and Objectives 11

Discussion 11

Section 3: Data source and Exploration *11*

REFIT House Data 11

Summary of data collection [1] 11

Timeline [1] 12

Data statistics [1] 12

Section 4: Machine Learning *12*

Getting started 12

Picking the data for machine learning 12

Preparing the Data for Machine Learning 13

Comparing Machine learning techniques 14

Testing Random Forest Tree algorithm with Under-Sampled Data 14

Section 5: Results and Discussion *15*

Conclusion *20*

Appendices *21*

1. **Figure 6 Breakdown for individual houses** 22
2. **Figure 8 breakdown for individual houses** 23
3. **Visualisation of predictions** 24
4. **Timeline for the Project** 24
5. **Dataset Details** 24

Acknowledgements 24

REFIT: Electrical Load Measurements [20] 24

Exploring the Data 25

6. **Random forest Regression** 25

Bibliography *25*

Table of Figures

Figure 1 visual representation of the idea 10
Figure 2 Visualisation of on/off data created from electrical consumption 14
Figure 3 Performance confidence interval for houses 2-4 with Kettle, Microwave, TV; electrical load measurement data resampled at 5 minutes; Method was rerun 200 times 16
Figure 4 Performance confidence intervals for House 2; electrical load measurements resampled at 1 minute; method was rerun 100 times 16
Figure 5 Accuracy and Recall score averaged over 17 houses; RFT applied to under-sampled data; Load Measurement data resampled at 5-minute frequency 17
Figure 6 TV; Accuracy and Recall score averaged over 16 Houses; RFT applied to under sampled data; Load measurement data resampled at 1-minute frequency 17
Figure 7 TV on/off time vs predicted on/off time; House 5; 1-minute load measurement resampling 18
Figure 8 Microwave; Accuracy and Recall score averaged over 14 Houses; RFT applied to under sampled data; Load measurement data resampled at 1-minute frequency 18
Figure 9 Kettle; Accuracy and Recall score averaged over 12 Houses; RFT applied to under sampled data; Load measurement data resampled at 1-minute frequency 19
Figure 10 Kettle on/off time example 20
Figure 11 House 5; 5 minute interval resampling of the load measurement data 24
Figure 12 RFT regression applied to living room temperature from House 1 25

List of Tables

Table 1 Comparison of most popular smart home interfaces 8
Table 2 Key Figures of the Dataset 12
Table 3 Data structure of electrical load measurements 13
Table 4 Data structure of binary on-off table for the appliances, after breaking up the index 13
Table 5 comparison of machine learning techniques in House 2 TV 15
Table 6 comparison of machine learning techniques in House 2 Microwave 15
Table 7 Average feature importance for TV; Columns represent the houses with TV in them 18
Table 8 Average feature importance for Microwave; Columns represent the houses with Microwave in them 19
Table 9 Average feature importance for Kettle; Columns represent the houses with Kettle in them 19
Table 10 Mean feature importance of appliances 19

Adaptive Function grouping in Smart Homes

The Abstract

The purpose of the report is to demonstrate that user behaviour in homes can be learned and predicted based on time of the day. The report makes a case for how this functionality would improve on existing home automation user interfaces and how it would improve the user experience. It identifies that data

about behaviour in homes is often unbalanced as a lot of things are mainly in standby, so the datasets have to be rebalanced before machine learning is applied. In the report the user behaviour was simulated from the electrical load dataset contained in the REFIT public dataset [1]. In the end it was found that the most effective way to predict user behaviour based on time is to rebalance the dataset by under sampling it, creating a binary classifier for each appliances' on-off state based on its electrical load pattern, then applying Random Forest Tree classifier algorithm using the features: month, week, day of the week, hour, minute. This approach yielded over 80% accuracy with multiple behaviours patterns across 20 UK homes.

The Introduction

Home automation encompasses the remote and centralised control of, heating, cooling, security, entertainment and lightning in homes.

The home automation market is in its early growth stage right now as no dominant player has emerged yet and it is just on the verge of breaking into the mainstream consciousness. Yet it is not something to be ignored as its value is expected to rise and it has gained the attention of some of the biggest tech companies like Apple, Google and Amazon.

The control platforms provided by mentioned companies have sparked an upheaval in start-ups creating products compatible with them, from locks to smart thermostats. These efforts have created the lower end of the home automation market by creating solutions that are easy to set up or retrofit. The high end of the home automation market consisting of integrated solutions, that are built into the house, has existed for several decades now, but it was only affordable for people at and above higher middle-class incomes and its adoption rate was low due to poor user interfaces. In recent years the cost of entry to the high-end market has significantly decreased, but general consumers are still unaware of this.

Both high- and low-end solutions offer smartphone-based controls and touch interfaces. Currently the high-end market does a better job at integrating all controls into a single interface, as some of the lower end solutions still rely on individual smartphone apps which don't communicate between each other, but Apple, Google and Amazon are all pushing developers towards integrating their solutions with their platforms. (Google Assistant [2], Amazon Alexa [3], Apple HomeKit [4]).

The fact that such big companies are heavily investing in the market and their approach right now is creating a better user interface (HomeKit for Apple, Alexa for Amazon and Google Home for Google), suggests that improvement is needed in that area and the main challenge for wider adoption is indeed lacking user experience right now.

This report is going to examine the main challenges for creating, a compelling smart-home interface. And explore a possible solution, which involves user behaviour prediction to bring forward functions that are expected to be used during a given time of the day.

This idea was developed based on feedback collected from companies working in the smart-home industry and in person user interviews, to find the biggest pain points of customers. [5] [6]

Because the home automation market is so scattered and it is hard to navigate compatibility, the user interface was approached from a software perspective, looking at the main inputs any system might get from a home and how to translate that into useful information. At the current state of smart-home interfaces there are multiple competing solutions, but they are all built around the same principles. Offering remote control of the connected devices and possibility of setting up scenarios, which are a series of actions triggered by a sensor or user input, that take care of multiple things in the home. The scenario set up ranges from not intuitive to very difficult. In systems with high integration an engineer is required to do it, and in the mentioned tech companies' ecosystems the user has to input all the commands manually which goes beyond the patience and ability of most customers.

The current approach to creating scenarios is lacking because it is time consuming and requires users to know the exact way, they intend to use their home and their habits, about which they are often wrong. Also once set these scenarios are not easily adaptable.

All these issues hinder every current smart home solution from becoming a seamless experience resending into the background while offering robust automation for the user.

A system that learns from the behaviour of the user and is able to predict usage based on time would be able to contextually group together functions in the home based on the time of the day, effectively creating scenarios for the user based on the time of the day, while it continuously learns the user's preferences.

This report will explore the above proposed solution and how it would improve home automation control interfaces. Starting with looking at current solutions then, at the findings in person interviews and research.

Introducing the goals of the project from the above findings the report continues with the description of the dataset that was used for simulating behaviour of the home occupants. It describes the decision process of picking the right measurements from the dataset and continues with the preparation of the data for machine learning algorithms.

Logistic Regression, KNN are applied to the dataset and from the results it leads onto why a more robust approach is needed as well as explores how the unbalance of the dataset (mainly consisting of zeros) messes with the model fitting of the algorithms.

Further two classification algorithms are explored: Bagging classifier and Random Forest Tree classifier (RFT), and under sampling and oversampling methods to rebalance the dataset. In the next section of the report the results are presented from the described processes along with important tables and figures and the results are discussed. In the conclusion the results are summarised, and a conclusion drown from the findings along with suggestions for the implementation of the method for behaviour prediction that was found during the experiment.

Section 1: Research and Preparation

Literature review

Home automation technology is yet to gain mainstream traction despite being around for the past 4 decades. The main hurdles that stand in the way of wider spread adoption are mostly the same: High cost of ownership, inflexibility, poor manageability and security [7].

Hardware

These challenges are slowly addressed by different approaches, it is yet to be seen which ends up being the most successful. The cost of ownership is declining each year, as the sensors and components get cheaper and wireless solutions are worked out. Internet of things protocols are getting more reliable and robust and this led to the development of low power reliable wireless networks (for example ZigBee and Z-Wave [8]). These wireless standards are being developed for reliable indoor use which would reduce the need for wiring significantly for complex home automation. This would mean complex home automation could be implemented in homes without the need for construction. This brings down costs significantly (here is an example of smart home application implemented with ZigBee instead of wiring [9]).

Many home automation systems suffer from compatibility issues and limit customers choice of new hardware in the home. Manufacturers like to develop their own standards, as that makes compatibility with their own hardware easier and locks users into using their products (good example of this is Apple). A study conducted on novel technology development in the 90's concluded that closed ecosystems and companies with narrow technological focus are not ideal for developing home automation solutions [10].

The home automation market is really fragmented due to the above-mentioned tendency, and it is a complex job navigating which products are compatible. But there is progress towards increased compatibility. Currently the most supported international standard is KNX [11] which is a device to device communication standard that is used at the core level of the system, and companies build different user interfaces around it. It is supported by almost all major component manufacturer, enabling a new level of flexibility in smart-home design allowing devices from multiple manufacturers to work together and be controlled from the same central control unit. But there are still manufacturers that close off their control systems built onto KNX and only their solutions communicate through the standard for example, Bticino, Creston, Bang and Olufsen.

KNX standard is a good base for a system as almost all new standards have gateways developed to connect to it. [12] [13]

Currently the hardware options can be very confusing as the earlier described high-end and low-end market start to merge or meet in the middle. Integrated solutions built around KNX standards can now be connected with Google, Apple and Amazon control interfaces, introducing voice control features and tying the control interface of the home into one of the apps developed by the 3 companies. [14] [15] Currently support for connecting to any of the mentioned platforms is software based. Apple switched to this approach only 2 years ago up until then they required a specific chip in the device to make it compatible with their system.

Hardware that is compatible with Amazon, Google or Apple platforms comes from a very wide range of manufacturers and it is becoming increasing common practice to support multiple platforms on a single device, which means the differentiating factor for the platforms in the future will be the quality of the user interface and not compatibility restrictions, if current trends continue.

These standards are all local networks and their connection to the outside world is through the WIFI internet network of the home. Thus, for increased security in smart homes the most obvious step is increasing the security of home routers. Wireless protocols are evolving in terms of security but, they are still less secure than their wired counterparts.

User Interface and Software

The last main limiting factor mentioned was poor manageability. Based on the hardware side of things the industry has reached a point where it has the technology to offer a great user experience. But poor manageability is mainly due to the lack of a great user interface.

User interfaces certainly have improved over the years, thanks to better touchscreen displays and cheaper processing power the presentation of the controls have become smoother and visually more appealing. But the user experience still hasn't improved greatly as the approach to user interface design stayed the same. Which is give more control to the user, but the more connected devices there are the more overwhelming and tiring it gets for users. Currently there are very few solutions on the market that leverage data to simplify the controls of the home and they are focused only on one aspect (Hive, Nest thermostats are an example).

The current systems on the market only react to predefined sensor readings or execute predefined scenarios. There is currently research going into user behaviour analyses and improvement of user experience. This is more challenging than traditional market segments due to the high number of variables in smart home designs and use cases. In [7] a comprehensive study with 14 participants living in smart homes found that while most homes had similar features and general capabilities the way this was achieved varied greatly and the potential for further automation was vastly different between homes.

The common features focus around energy, money saving and time saving as the top priorities based on market research of current systems. Another common trend is that they offer touchscreen and smartphone controls as the main methods of interactions and their features are fixed [16].

These interfaces in integrated systems are usually built onto the control interface provided by the device manufacturers. Common practice is when a home automation company creates a solution for a client, they build a costume interface for them. When it is a do-it-yourself installation the control interface consists of separate apps or ties into a one of the 3 mentioned tech companies' platforms.

Control interface comparison

Company	Apple	Google	Amazon	Schneider Wiser	Boticino	Lutron	Creston	Control4	Bang and Olufsen
Smart phone controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
voice control	yes	yes	yes	no	no	no	no	no	no
costume interface	no	no	no	yes	yes	no	yes	no	partially
user configurable interface	yes	yes	yes	no	no	no	no	no	no
adaptive interface	partially	partially	no	no	no	no	no	no	no
scenario creation	yes	yes	yes	yes	yes	yes	yes	yes	yes
scenarion creation by customer	yes	yes	yes	no	no	no	no	no	no
energy consumption monitoring feature	yes	not sure	not sure	yes	yes	no	no	yes	not sure
learning features	limited	limited	limited	no	no	no	no	no	no
remote acces	yes	yes	yes	yes	yes	yes	yes	yes	yes
security level	high	medium	medium	high/medium	high	high	high	high	high
processes data locally	yes	no	no	yes	yes	yes	yes	yes	yes
needs cloud connection	only for voice	yes	yes	no	no	no	no	no	no
main interface	touch + voice	voice	voice	touch + switches	touch + switches	touch + switches	touch + switches	touch + switches	touch + remotes

Table 1 Comparison of most popular smart home interfaces

Table 1 is a summary of the key points of some of the most common user interface providers at the moment. The 2 colours separate the big tech companies from the rest of the options as fundamentally their approach differs in one main thing: Apple Google and Amazon (AGA) [17] make it accessible to casual users to add devices to their platform and configure scenarios, rooms and controls, the system can reach a high automation level before the customer needs to turn to a specialised installer and most of their systems are retrofittable. Their main advantage lies in this and the fact that their software looks and feels the most polished as they have more expertise and resources in building user interfaces than industrial companies or boutique automation companies.

The companies in blue are building costume user experiences and they work with authorised installers in each country. Their systems have to be planned and tailored to the users' needs and usually require a higher level of upfront investment. Most of them provide costume controllers built for their systems. Their main advantage compared to AGA is that they support a much wider range of devices that can be controlled through their interfaces and gateways. They all support the KNX standard mentioned earlier to communicate with a wide range of appliances and devices in the home. These systems need to be installed during construction or remodelling of the home as they build onto wired and wireless communications as well and are intended as a comprehensive all house solution. Their main drawback lies in the fact that the user experience is very reliant on the installers understanding of the users' needs, as it can only be reconfigured by them. The connection between devices this way tend to be more stable and reliable though, but AGA certified devices are catching up quick in this department.

The user interface of companies in blue are better in bigger houses as they allow for function mapping to switches in rooms and location specific control panels. Offering robust local controls in each room without the need for a smartphone but allow for all house control through a phone as well and remote access. Where they lack compared to the yellow section is the voice control and user configurability of the systems.

This is an area I believe the 2 approaches are going to form a symbiosis. Currently there are few very young companies creating KNX modules that connect all KNX integrated devices and setups into Apple, Google and Amazon interfaces completely. Thinka [18] seems as the best solution so far in this space, but at the point of writing this report I haven't got the chance to play around with it in person only looked at online demonstrations. (Will update the online appendices, if I get the chance of trying it in April 2019).

This solution would augment the integrated systems significantly giving users access to a better and constantly updated mobile user interface where they can configure further automatization and link their phone calendar and location to functions while leaving the costume interfaces of the original systems intact and fully functioning.

The natural evolution of the market seems that small retrofitted projects will solely rely on AGA platforms, and highly integrated systems will still have a place in the luxury market and the newbuilt market, but they will integrate support for AGA control platforms for Voice control and location-based services to reduce the need for extra sensors.

Based on these trends the best point in the system to introduce a learning algorithm to improve the adaptability of the system is at the controller level, were usage of things can be instantly saved digitally without the need for extra sensors. The traditional interfaces would have to do this on a local server where they store the data. The AGA platforms could leverage cloud computing and on device processing to serve up suggestions based on previous usage. They platforms are likely to have this type of functionality within 4-7 years' time as there are traces showing of such learning behaviour in their mobile operating systems.

Apples' Siri has started serving up app suggestions and actions like message or call a certain person, or open a document, based on location and time of the day. I believe this functionality will eventually make its way into HomeKit as well. Google has started doing similar time and behaviour-based suggestions in Google Maps and Gmail.

Based on the current developments of smart home user interfaces the next natural step in their evolution seems to be self-learning and the introduction of behaviour-based user suggestions as the cost of hardware is going down more and more things can be made smart in homes. Connecting more devices into the control interface complicates it further, all the mentioned interfaces suffer from over cluttering and they are at their best when scenarios are configured in them. Time based behaviour prediction would be able to serve up functions grouped together or just bring forward menu points based on time of the day, streamlining the user interface.

Additional research

Gathered additional information by visiting companies working on home automation. LifeEvolution [5] a home and building automation company based in Romania, got to inspect the technologies installed in their showroom and see how they design a home and got to visit one of their installations. Another company I got to visit was Elektra S.A. home automation department [6] and their showroom. Apart from this I visited a home where Control4 system was installed and got to speak to the users of it. [19]

After talking to the CEO's of both companies I came away with the following conclusions. Technologically right now a very high level of automation and comfort is achievable given the user has a clear image of their usage of the home (which they usually don't), but all these configurations have to be made by the installer which is time and money consuming. When asked about the challenges of the business they both brought up the lack of knowledge of costumers about the possibilities. Both said partnering with architects would be helpful as these technologies are often not considered when designing a home simply due to lack of knowledge and integrating the infrastructure for it during building is minimal cost compared to the construction cost.

When asked if a more intelligently adaptive service would be something, they find useful they both said yes. My takeaway from these conversations was the following:

- There is no fully integrated system that adapts to users on its own without an engineer reconfiguring it. This either has to be on site or remotely.
- The users are presented with data about their behaviour, but more often than not they don't tend to analyse it let alone act on it.
- Features that are buried in layers of menus tend to go unused even if they offer helpful controls.

Target Audience of solution

The target audience of the solution would be middle class and above homeowners. This interface improvement brings the most value in houses which have smart features in at least 3 to 4 rooms. Hence the assumption of target audience for the solution being middle class and above, they can afford to live in a home with several rooms. Home automation is now available on basic levels to almost anyone as its cost is low compared to the living spaces cost.

Section 2: Goals and Objectives

Having set out to create a better user interface for the smart homes, the following interface was envisioned, see mock-up below.

Figure 1 visual representation of the idea

The description of the proposed idea

The system would learn about the needs of the user and create scenarios based on the features used frequently together.

An example of this working presented through a generic morning routine:

The morning routine:

- 7:00-7:30: user wakes up
- After waking up uses bathroom for 20 minutes
- 8:00 in the Kitchen having coffee
- 8:30 leaving for work

The system after learning time, would present the following option to the user in the previous evening: Morning routine (bedroom blinds open at 7:00, Bathroom temperature set to 22 C from 7:00 till 8:00 and lights turn on, Kitchen lights turn on at 8:00 till 8:30, The coffee machine turns on at 7:50 so its preheated and ready to use by the time the user gets to the kitchen)

The underlying idea is to maximize the comfort of homes by reducing the number of times users are required to fiddle with the settings. Current systems are focused on giving more and more control to the user, but most of these features end up going unused due to lack of time and learning commitment required on the side of the users.

Goals and Objectives

1. Data processing. A general way to draw meaningful conclusions from sensor data in home and understand usage better. Identify the least intrusive and easy to set up sensor set to get useful data. This would address the need for better optimisation of the systems weeks after installation and would require less interactions between the user and the company and it would be more accurate, providing a better customer experience.
    a. Find or produce a dataset that is a good reflection of average home usage in the UK or Europe and is spanning at least a few months.
    b. Identify the things worth analysing from usage data.
2. Show that there are patterns in the user behaviour and there is correlation between time of the day, and day of the week and month of the year between habits and usage of functions.
3. Find an algorithm or process that can predict with 80% plus accuracy the behaviour in a house based from the given input data.
4. Create a program that groups functions together based on time of the day and rooms.
    a. Based on the results from the algorithm discuss where in the ecosystem can the solution be best integrated. Sensor level, central control unit, or on the controller itself.
5. Draw conclusions for improvement, and identify, next step on how to make the learning faster and more robust.

Discussion

The objectives listed set the steps for the project to give a structure of how the proposed solutions viability is intended to be explored. The main goals of the project are numbered while objectives related to them are below them in alphabetical order.

The Goals listed are things that can be useful for companies that install and develop home automation systems. Also, they build onto each other. By reaching the 3rd goal the viability of the envisioned system can already be proven as once there is an algorithm that can predict the usage pattern based on time only, the rest of the system only requires a time-based look up table.

Section 3: Data source and Exploration

After looking for national statistics on home usage I came across a very useful dataset and building onto that I could start progressing my project.

REFIT House Data

The dataset creation was led by Dr Firth who has been contacted, and he agreed to the use of the dataset for this project as long as it is correctly cited. [1]

The official description of the data follows below.

Summary of data collection [1]

The dataset is for 20 homes located near to the town of Loughborough in the East Midlands region of the UK.

A building survey was carried out at each home, collecting data on building geometry, construction materials, occupancy and energy services.

Each home has a selection of the following sensors and devices installed:

- CurrentCost mains clamps, to measure household mains electrical power load
- Replacement gas meters, to measure household mains gas consumption.
- Hobo pendant or Hobo U12 sensors to measure room air temperature, relative humidity and light level.

- iButton temperature sensors to measure radiator surface temperature.
- CurrentCost individual appliance monitors, to measure plug electrical power loads
- RWE Smart Home devices including programmable thermostatic radiator valves, interior and exterior motion detectors, door and window opening sensors and smoke alarms.
- British Gas Hive programmable thermostats.

In addition, climate data was collected at the Loughborough University campus weather station.

Timeline [1]

September 2013 to February 2014: Building surveys were carried out and monitoring sensors were placed in the buildings at or shortly after this time.

June 2014 and October 2014: Smart Home devices were installed in the buildings.

April 2015: Data collection finished.

Data statistics [1]

Dataset key figures	
Number of homes:	20
Number of spaces (rooms):	389
Number of radiators:	252
Number of showers:	34
Number of appliances:	618
Number of light bulbs:	672
Number of fixed heaters:	19
Number of surfaces:	2237
Number of openings:	970
Number of sensors:	1567
Number of variables recorded by sensors and devices:	2457
Number of time series readings:	25312397

Table 2 Key Figures of the Dataset

Section 4: Machine Learning

Getting started

Due to the size of the datasets [20] [1] the project had to be completed in Python as MATLAB was not suited to deal with the size of the CSV files that contained the measurement points. MATLAB was used to explore the structure of the xml file, by creating a struct variable from the xml that made it easier to explore how the dataset was composed and what it contained.

The struct file created in MATLAB was used to extract paths towards variables in the dataset, which were then linked to the csv file in Python allowing the data to be visualised and plotted.

Picking the data for machine learning

Python libraries used at this stage: NumPy, Pandas.

The data from motion sensors was analysed and it was found that it is quite inconsistent and unreliable within the dataset. By examining the air temperature and quality in the rooms it was found that a cyclical behaviour based on occupancy of the rooms can be shown, but it was also influenced by the outside temperature and the amount of solar radiation hitting the side of the house where the rooms were. Based on the mentioned factors these data points were ruled out. But as a side result a strong correlation between gas consumption indoor and outdoor temperature could be shown and predictions made about either based on the other with a few regression algorithms. The same relationship turned out to be true for inside air temperature and relative humidity and light intensity.

From the light intensity and the outside solar radiation relationship it could be deducted if lights were on or off in a room, but that didn't give enough behavioural information about the occupants.

All the mentioned approaches had one other limiting factor in common, the measurements were recorded at 30-minute intervals which doesn't allow for a down to minute level definition of user behaviour.

By looking at the electrical consumption data it was found that it is quite easy to deduct the operating interval for most appliances, and from that to tell when somebody is using them. As every home had 9 appliances recorded this would serve as a good basis to simulate the usage of a home as based on the appliance it can be categorised which room is being used and for how long, and how many rooms are being used at the same time. Additionally, these measurements were saved at every 5-8 second intervals, which gave a better time breakdown of behaviour.

To deduct this usage data, it was necessary to exam the usage patterns of these appliances as some are only attended by a user at the start and end of operation for example a dryer or a washing machine. But each house contained at least a couple of appliances which would have continues consumption while actively in use by someone and drop sharply when off. Most notably every household except one had a TV that was monitored. Other examples that were quite common and easy to monitor were toaster, kettle, microwave, computer, Hifi these are appliances that usually all require the user's presence so are ideal for the study and have a consumption pattern that is easy to filter for on and off behaviour. The appliance that was present in every home was fridge and freezer, but the consumption pattern is very hard to analyse for door openings [21] which would indicate user interactions, so these were ignored after exploring a bit the possibility of including it.

By looking at the findings of the analyses of the different parts of the dataset it was found that the electrical load measurements dataset is the best to be used for behaviour simulation.

Preparing the Data for Machine Learning

For each house a usage data frame was created. The data frame would contain a binary classifier for each appliance indicating if it was on or off, the index was set to the date time index of the electrical load measurement dataset. Then based on the load pattern the data frame was filled in with the right information. The load measurement dataset was resampled with 5-minute intervals before the classification was done to reduce the size of the data set to a size that is less computationally tasking so the programs run time stayed reasonable (Table 3).

Time	aggregate	fridge	washing_machine	dishwasher	TV	Microwave	Toaster	Hifi	Kettle	Fan	issues
2013-09-17 22:05:00	670.0588235	87.64705882	0	0	0	0	0	0	0	0	0
2013-09-17 22:10:00	729.7333333	86.17777778	0	0	0	0	0	0	0	0	0
2013-09-17 22:15:00	718.94	84.36	0	0	0	0	0	0	0	0	0
2013-09-17 22:20:00	621.12	82.54	0	0	0	0	0	0	0	0	0
2013-09-17 22:25:00	166.74	59.12	0	0	0	0	0	0	0	0	0

Table 3 Data structure of electrical load measurements

It is not advised to run machine learning on time series data as the algorithm will struggle to find patterns in date formatted data. Because of this the index was broken down into month, week, day of the week, hour, minute. These are ideal features as with enough data the model could potentially distinguish between days of the week and times of the year, not just the time of the day (Table 4).

Time	month	week	day	hour	minute	aggregate	fridge	washing_machine	dishwasher	TV	Microwave	Toaster	Hifi	Kettle	Fan	issues
2013-09-17 22:05:00	9	38	1	22	5	0	0	0	0	0	0	0	0	0	0	0
2013-09-17 22:10:00	9	38	1	22	10	0	0	0	0	0	0	0	0	0	0	0
2013-09-17 22:15:00	9	38	1	22	15	0	0	0	0	0	0	0	0	0	0	0
2013-09-17 22:20:00	9	38	1	22	20	0	0	0	0	0	0	0	0	0	0	0
2013-09-17 22:25:00	9	38	1	22	25	0	0	0	0	0	0	0	0	0	0	0

Table 4 Data structure of binary on-off table for the appliances, after breaking up the index

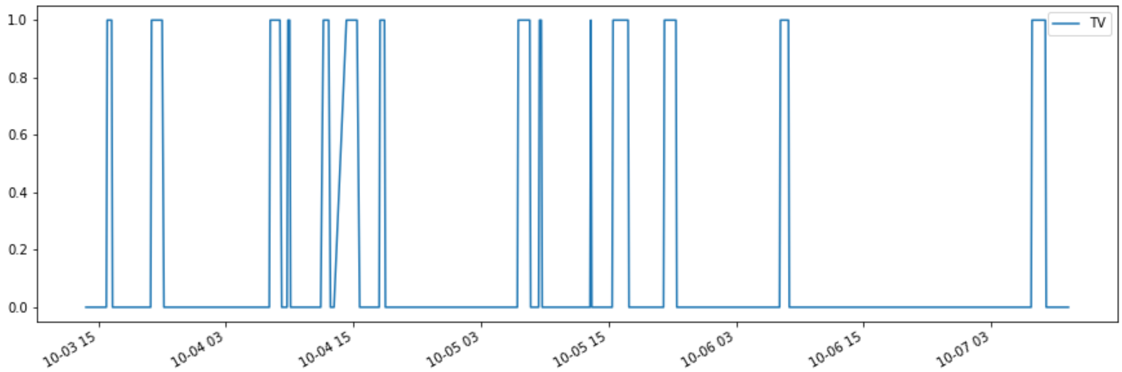


Figure 2 Visualisation of on/off data created from electrical consumption

Once the usage table was created, the rows which contained potentially faulty data which is marked by the issue's column were dropped.

Comparing Machine learning techniques

All the machine learning algorithms were imported from the scikit-learn library in python. The following classifiers were applied to the data: Logistic Regression [22], KNN [23], Bagging [24] and Random Forest Tree Classifier [25] [26].

KNN and Logistic regression both reached 90%+ accuracy scores. Bagging and Random Forest Tree (RFT) performed between 80-90%.

The result from this test all over fitted and predicted only zeros with the exception of RFT and Bagging. The issue was that the datasets created were predominantly made up of zeros because most appliances are not in use for the majority of their lifetime. The dataset contained between 85% - 98% zeros depending on the appliance.

The two approaches explored to deal with the unbalanced dataset were under-sampling and over sampling. Under sampling consists of reducing the number of the entries that are in majority by making it equal to the other entry. In practice this meant saving the number of ones, then randomly picking equal number of zeros from the zero indices and creating a new balanced dataset this way. Over sampling meant introducing more ones in intervals were ones are present until the number of one's equal the number of zeros, this meant artificially increasing the sampling rate during periods when the appliance is on.

The two main metrics to assess the performance of the models was accuracy score and recall score. Accuracy score is the percentage of accurately predicted instances. The recall score is the percentage of the correctly predicted ones. High accuracy and low recall score indicate that the model predicts too many zeros. The ideal outcome is a high accuracy and high recall score.

After running all the algorithms again with the under-sampled data and the oversampled data it was apparent from the two metrics mentioned that the best results were yielded by applying the Random Forest Tree classifier to under-sampled data (Table 6).

Testing Random Forest Tree algorithm with Under-Sampled Data

The differences in performance proportionally stayed the same when the methods were applied to different appliances.

The best results were yielded consistently by under sampling and Random Forest Tree classifier combination, it was picked for further testing. After running it multiple times on the same dataset a small deviation was noticed in this methods performance. This is due to the randomness of the under-sampling of the data. As the zeros that are kept in the dataset are selected at random, sometimes the randomly removed zeros would have contributed to a better model.

To assess the effect of randomness in under sampling on the average performance of the method the under-sampling and model fitting was rerun for several appliances in different houses 200 times, saving the recall and accuracy each time. From the saved accuracy dataset, a distribution graph was created, which reassembled that of a gaussian distribution, and gave the confidence interval for the method.

To prove that the method was applicable in any arbitrary instance the performance of it was tested across multiple houses. An appliance was picked that was present in most houses. For each house the data was under sampled for the appliance, then the random forest tree was fitted to it with 75 times each time changing the splitting of the dataset into testing and training sets. Creating a dataset of how the model performs with 1-75% of data used for training. After this was done for each house the results were averaged for each percentage creating a good general indicator of how this method would perform in an arbitrary home.

Section 5: Results and Discussion

Untouched Data				
	KNN	Logistic Regression	Random Forest Tree	Bagging
Accuracy score	0.93244854	0.9322235	0.91049375	0.9129456
Recall score	0.02090939	0	0.27812811	0.3460294
Under Sampled Data				
	KNN	Logistic Regression	Random Forest Tree	Bagging
Accuracy score	0.7027516	0.547701	0.86356895	0.855165
Recall score	0.77748	0.557264	0.911321989	0.844785
Over Sampled Data				
	KNN	Logistic Regression	Random Forest Tree	Bagging
Accuracy score	0.74830727	0.5396276	0.84381959	0.8432572
Recall score	0.708596	0.522905	0.54530368	0.228178

Table 5 comparison of machine learning techniques in House 2 TV

Untouched Data				
	KNN	Logistic Regression	Random Forest Tree	Bagging
Accuracy score	0.99323	0.99323	0.990741	0.9907094
Recall score	0	0	0.084905	0.094339
Under Sampled Data				
	KNN	Logistic Regression	Random Forest Tree	Bagging
Accuracy score	0.673076	0.584134	0.8028846	0.795673
Recall score	0.812182	0.63959	0.80203	0.761421
Over Sampled Data				
	KNN	Logistic Regression	Random Forest Tree	Bagging
Accuracy score	0.883149	0.591692	0.9612732	0.958782
Recall score	0.5188679	0.589622	0.419811	0.424528

Table 6 comparison of machine learning techniques in House 2 Microwave

The above performance metrics (Table 5 Table 6) were the ones that led to the decision of sticking with the under-sampled data and fitting the model to it with Random Forest Tree classifier.

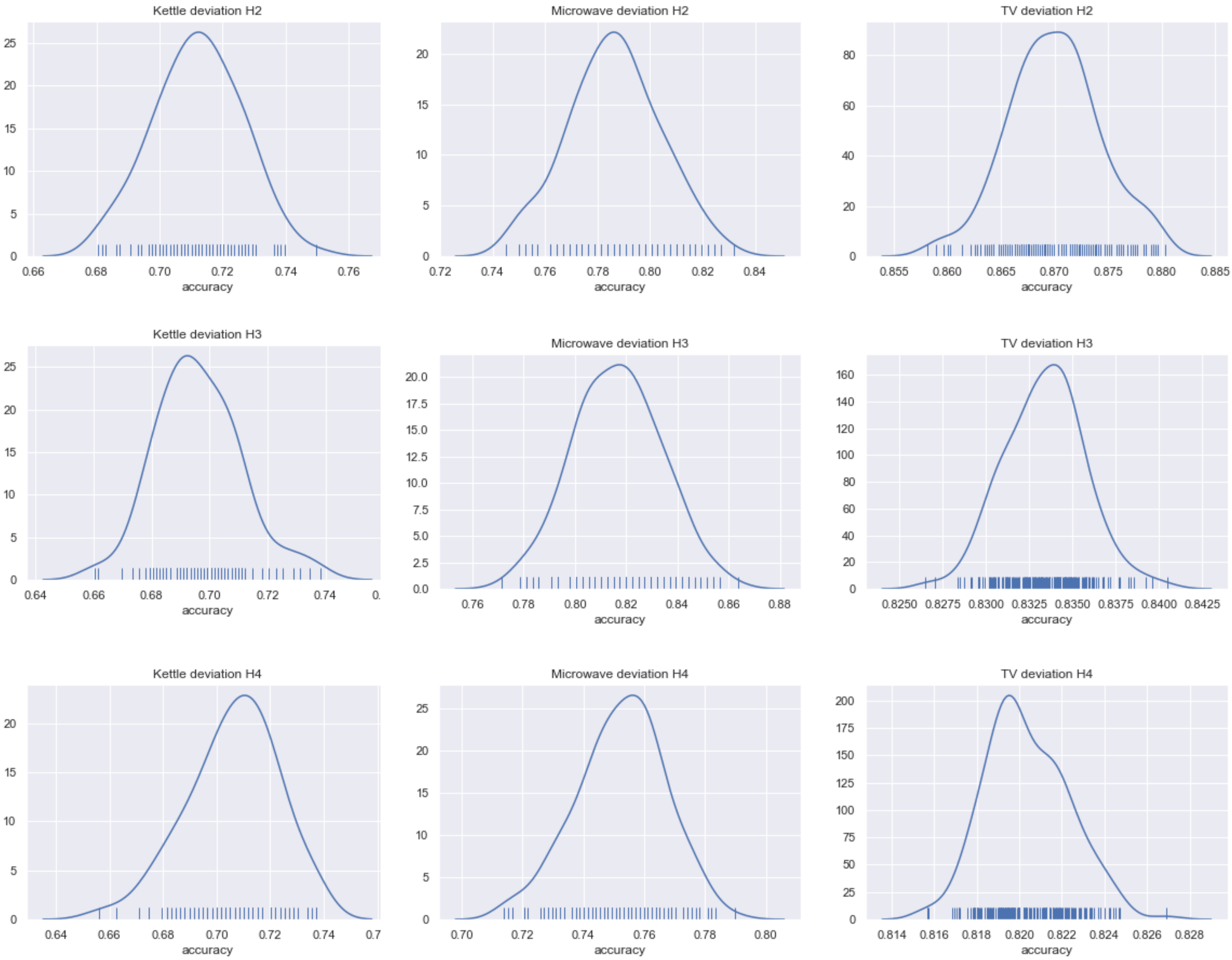


Figure 3 Performance confidence interval for houses 2-4 with Kettle, Microwave, TV; electrical load measurement data resampled at 5 minutes; Method was rerun 200 times

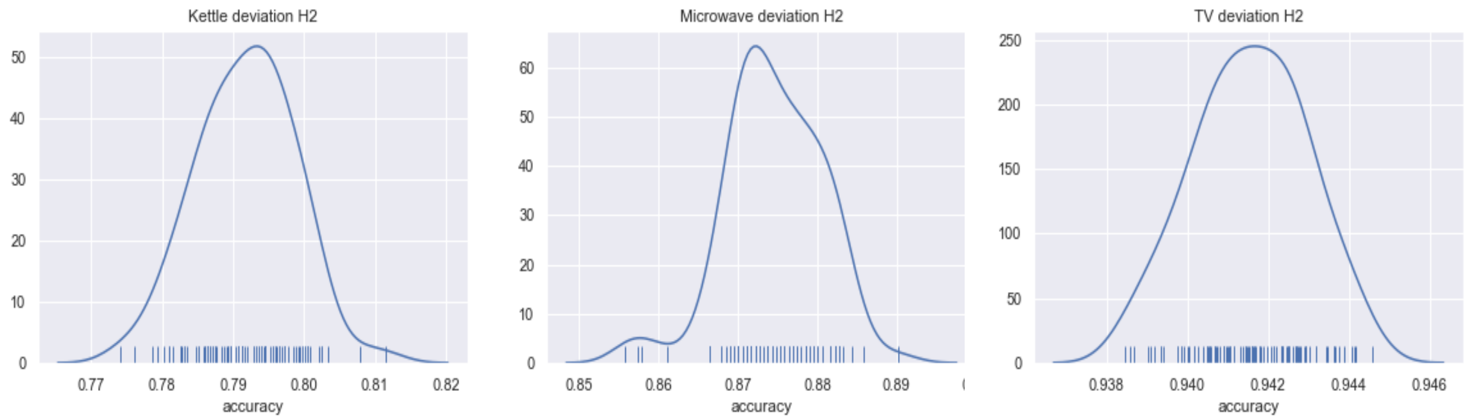


Figure 4 Performance confidence intervals for House 2; electrical load measurements resampled at 1 minute; method was rerun 100 times

The confidence interval closely reassembles a gaussian distribution because its randomness is due to the randomness of the under-sampling. From the graphs it can be observed that the accuracy stays more consistent with things that are on for longer periods of time (Figure 3). As in each case the TV 's deviation is the smallest as its generally used for longer intervals, while the Kettle and Microwave is only used for 5 to 10-minute intervals usually, sometimes even less.

The same analyses were rerun on the appliances of house 2 but with electrical load data resampled at 1-minute intervals. As the above figures (Figure 4) show the performance of the method increases if a higher sampling is used, as well and the model accuracy becomes more stable, less influenced by the under-sampling.

This suggests that the frequency of the saved data points should be decided based on the average on time of the function or device in the home. This would mean for functions or things that are only used for very short periods of time, like a Kettle the sampling time of the usage should be high, while for things that are on for longer times a lower sampling rate is enough to fit a good model. Based on the on average on-time of lights in a home and heating, this method would probably yield very accurate predictions. For example, of TV on-off-times (see Figure 2). As the TV was the most common appliance in the homes, to validate the method described I've analysed the performance of it on predicting TV usage in the homes. Figure 5 shows how the mean accuracy and mean recall improves across all houses with the increase in the size of the training dataset. The whole dataset is 3 years long which means that with only 3 months of training data the algorithm is already around 82% accurate.

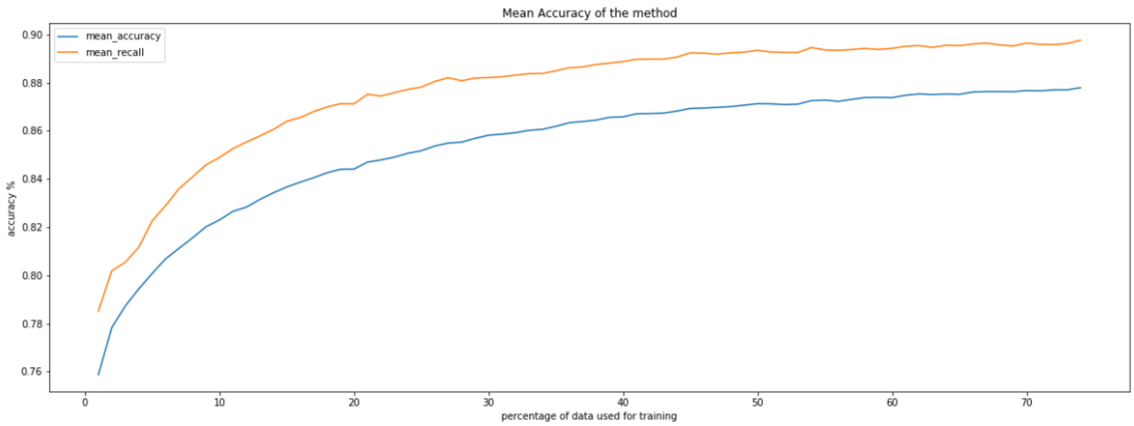


Figure 5 Accuracy and Recall score averaged over 17 houses; RFT applied to under-sampled data; Load Measurement data resampled at 5-minute frequency

The RFT classifier used on under-sampled data can serve as a good basis for an adaptive smart-home interface, as it reaches acceptable accuracy within a short period of time. The average accuracy can be even higher if a higher sampling rate is used for the data. Figure 5 was achieved with a sample rate of 5-minutes.

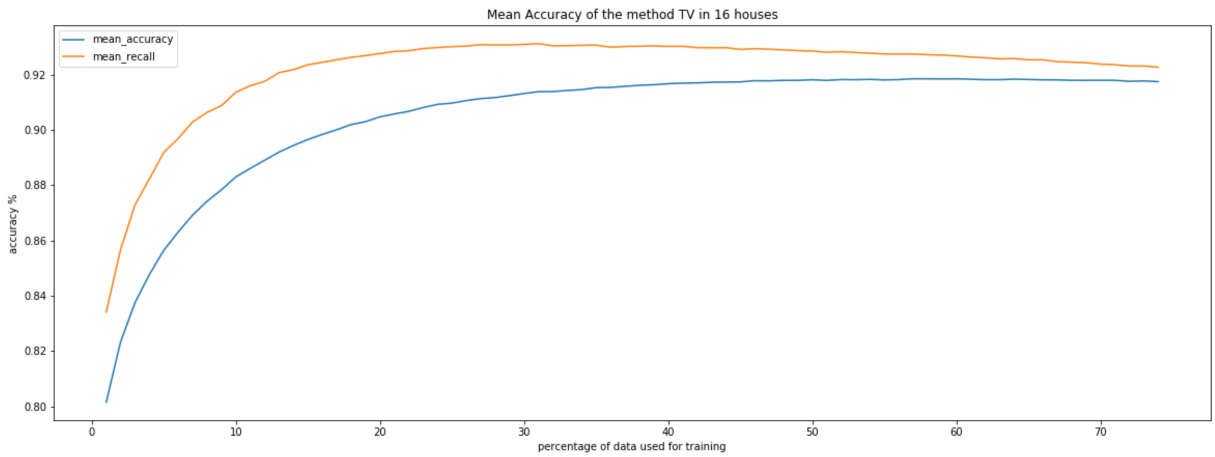


Figure 6 TV; Accuracy and Recall score averaged over 16 Houses; RFT applied to under sampled data; Load measurement data resampled at 1-minute frequency

Figure 6 shows the methods performance on TV again, when the electrical load measurement dataset is resampled at 1-minute frequency. The reason for dropping to 16 houses from 17 compared to the previous model is that running the tests for a 5 times bigger dataset significantly increases the time required and in 1 house the TV has different consumption threshold which couldn't be filtered out in a loop. It can be seen that increasing the sampling rate leads to a quicker learning curve.

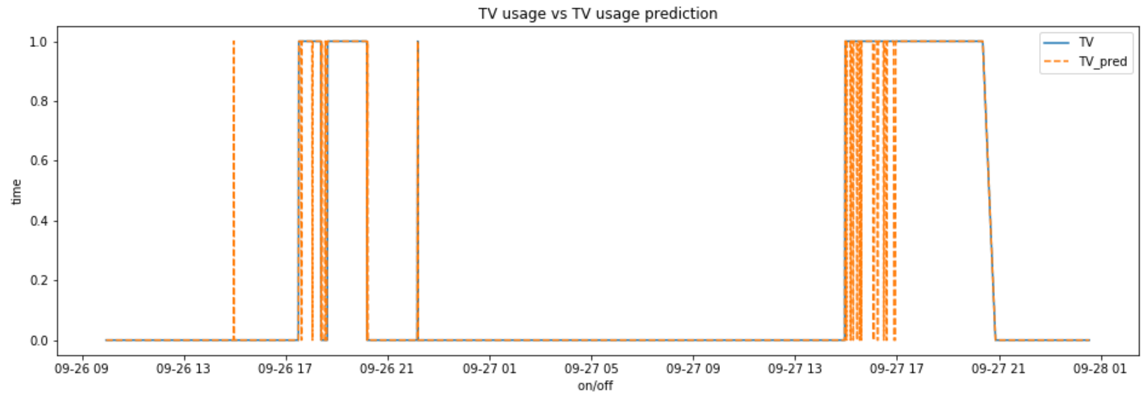


Figure 7 TV on/off time vs predicted on/off time; House 5; 1-minute load measurement resampling

The average feature importance for the TV was calculated by fitting the model to 75% of the data from each House (Table 7):

Average feature importance for TV																	
	mean	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**13**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
month	0.047332	0.04506	0.04838	0.029525	0.062831	0.048323	0.03402	0.031661	0.029988	0.033701	0.09259	0.035126	0.038391	0.062651	0.034495	0.038783	0.045553	0.093558
week	0.200284	0.233929	0.206664	0.181771	0.203429	0.204911	0.152371	0.144241	0.126201	0.160786	0.263401	0.179746	0.214406	0.269725	0.19562	0.206611	0.174406	0.286606
day	0.487228	0.385994	0.469036	0.463392	0.371003	0.48623	0.565966	0.572378	0.652622	0.615135	0.423841	0.535996	0.414416	0.396767	0.592796	0.45104	0.539213	0.34705
hour	0.14733	0.21414	0.171454	0.19412	0.232785	0.143247	0.145299	0.156116	0.115672	0.094941	0.060541	0.136537	0.189497	0.135091	0.092137	0.147903	0.130997	0.144137
minute	0.117826	0.120877	0.104466	0.131192	0.129952	0.117289	0.102344	0.095604	0.075516	0.095436	0.159627	0.112594	0.143289	0.135766	0.084952	0.155662	0.109831	0.128649

Table 7 Average feature importance for TV; Columns represent the houses with TV in them

The same test with 1-minute resampling was run on other common appliances in the homes.

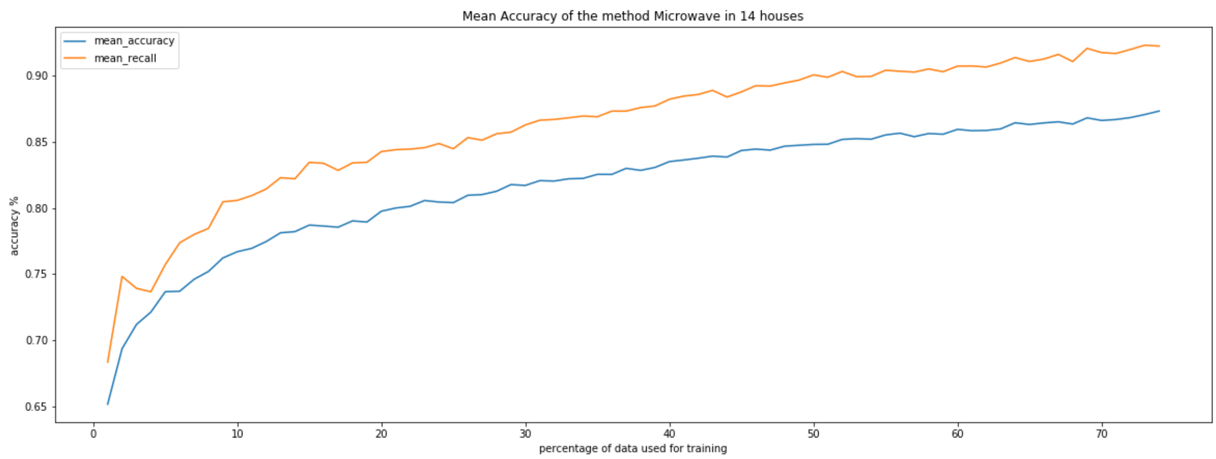


Figure 8 Microwave; Accuracy and Recall score averaged over 14 Houses; RFT applied to under sampled data; Load measurement data resampled at 1-minute frequency

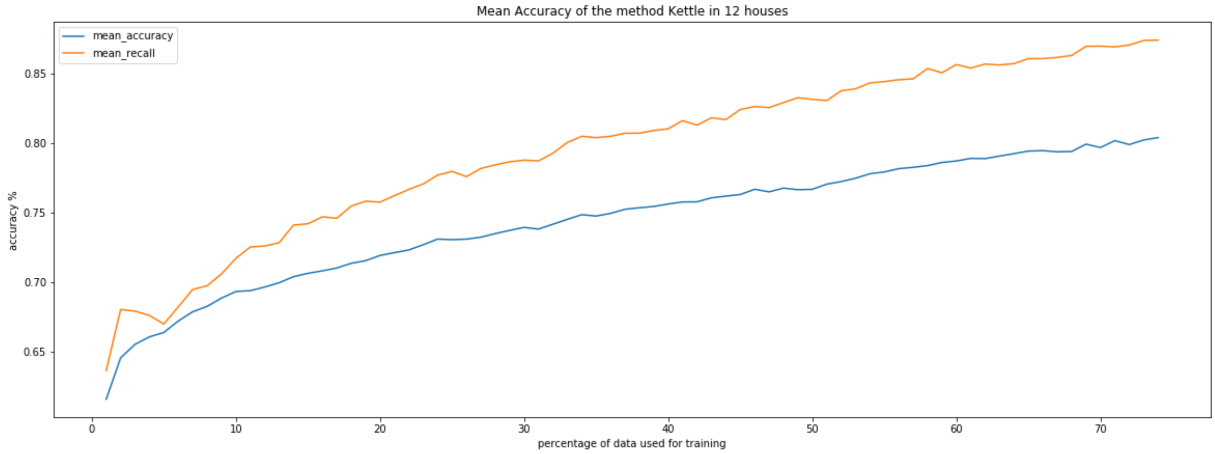


Figure 9 Kettle; Accuracy and Recall score averaged over 12 Houses; RFT applied to under sampled data; Load measurement data resampled at 1-minute frequency

Average feature importance for Microwave														
	mean	2	3	4	5	6	8	9	10	11	15	17	18	19	20
month	0.071178	0.116767	0.074182	0.055735	0.050208	0.049794	0.052469	0.068947	0.065701	0.103898	0.093569	0.066714	0.066144	0.066644	0.06572
week	0.17906	0.300187	0.172259	0.168902	0.165915	0.154892	0.145552	0.153387	0.175818	0.205091	0.199201	0.165366	0.138548	0.205784	0.155932
day	0.439311	0.264384	0.45443	0.414536	0.478169	0.455332	0.481125	0.527838	0.40335	0.364894	0.422183	0.451918	0.561386	0.429534	0.441277
hour	0.196962	0.184315	0.195304	0.230284	0.212755	0.221851	0.212033	0.139552	0.227942	0.195726	0.168892	0.203737	0.144113	0.191214	0.229748
minute	0.113489	0.134347	0.103825	0.130543	0.092953	0.118132	0.108821	0.110277	0.127189	0.130391	0.116155	0.112265	0.089808	0.106823	0.107323

Table 8 Average feature importance for Microwave; Columns represent the houses with Microwave in them

Average feature importance for Kettle													
	mean	2	3	4	5	6	7	8	9	17	19	20	21
month	0.059559	0.046903	0.052709	0.051933	0.054669	0.04634	0.058487	0.046796	0.053248	0.075378	0.077397	0.06812	0.082723
week	0.171415	0.173001	0.170653	0.158199	0.171389	0.157556	0.176662	0.163853	0.168098	0.188942	0.181673	0.164316	0.182639
day	0.396683	0.37566	0.379616	0.410065	0.396008	0.370044	0.421372	0.38678	0.389693	0.371914	0.422697	0.411232	0.425114
hour	0.260297	0.274383	0.284974	0.274122	0.279348	0.302891	0.237279	0.28993	0.285271	0.255075	0.199002	0.242802	0.198481
minute	0.112047	0.130054	0.112048	0.105681	0.098585	0.123168	0.1062	0.112642	0.103689	0.108691	0.11923	0.11353	0.111044

Table 9 Average feature importance for Kettle; Columns represent the houses with Kettle in them

Figure 8 and Figure 9 Show the second most common appliances after the TV. They are a good test for the performance of the algorithm because they are only on for short periods of times (see Figure 10). They reach acceptable level of accuracy which confirms the viability of the approach.

Mean feature importance values			
	TV	Microwave	Kettle
month	0.04733156	0.07117794	0.059559
week	0.20028389	0.17905961	0.171415
day	0.487228	0.43931125	0.396683
hour	0.14733023	0.19696183	0.260297
minute	0.11782632	0.11348937	0.112047

Table 10 Mean feature importance of appliances

Based on Table 7, Table 8 and Table 9 the feature importance of the model can be seen. Table 10 combines the mean values from the previous tables. It shows that each feature contributes meaningfully to the model, day having the highest importance in the prediction in each case. Interestingly for TV week is more important than the hour. But these findings suggest that behaviour patterns can be closely tied to days of the week and even weeks of the year. Feature importance could potentially be used to classify activities that are repeated each day, week or hour. These findings make sense as TV watching habits are expected to change considerably between days, assuming the user is employed.

Based on the above results it is feasible to assume that the method of creating binary usage data about behaviours could be used for learning the patterns in them. A lot of things in the home like lights, heating and presence in a room usually have longer active on-time than the appliances tested above. As the approach seems to increase in prediction accuracy the longer periods the observed behaviour happens, this leads to the assumption that the method could be effective at predicting the mentioned things as well.

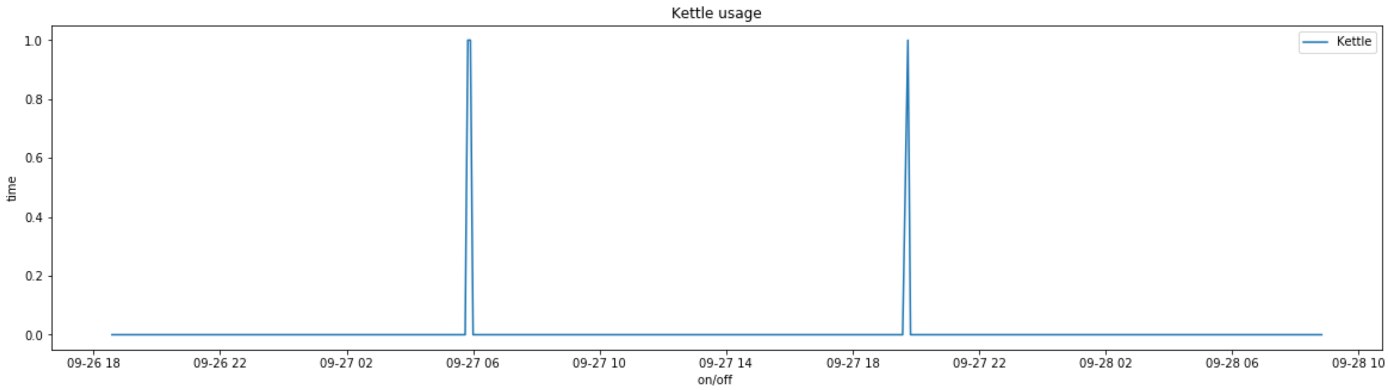


Figure 10 Kettle on/off time example

Based on the results the best point to integrate such solution into a system would be on a controller level, were every action can be saved digitally about all the appliances used, not requiring further sensor readings or installations in the home. If behaviour data is needed about a not-connected home, the findings of this report suggest that the easiest and least intrusive way to find out the most is by monitoring the electrical load of appliances in the home.

Conclusion

Looking at the state of home automation interfaces the next natural step seems to be time-based suggestions based on machine learning from previous behaviour. During the experiments the report describes it was demonstrated that time-based classification of behaviour is possible, and it can reach high accuracy in different usage patterns. The combination of balancing datasets by under-sampling and fitting Random Forest Tree classifier model to it yielded over 80% accuracy for predicting user behaviour correctly to 1-minute intervals. This level of accuracy would be enough for the proposed interface design in the beginning of the report, which aims to group together function that are likely to be used during the same hour and groups them together into scenarios for the user, simplifying the user interface significantly. The system could get even better if the feedback of user behaviour is introduced about how many times its suggestions were picked and which ones. But simulating that was outside the scope of this project.

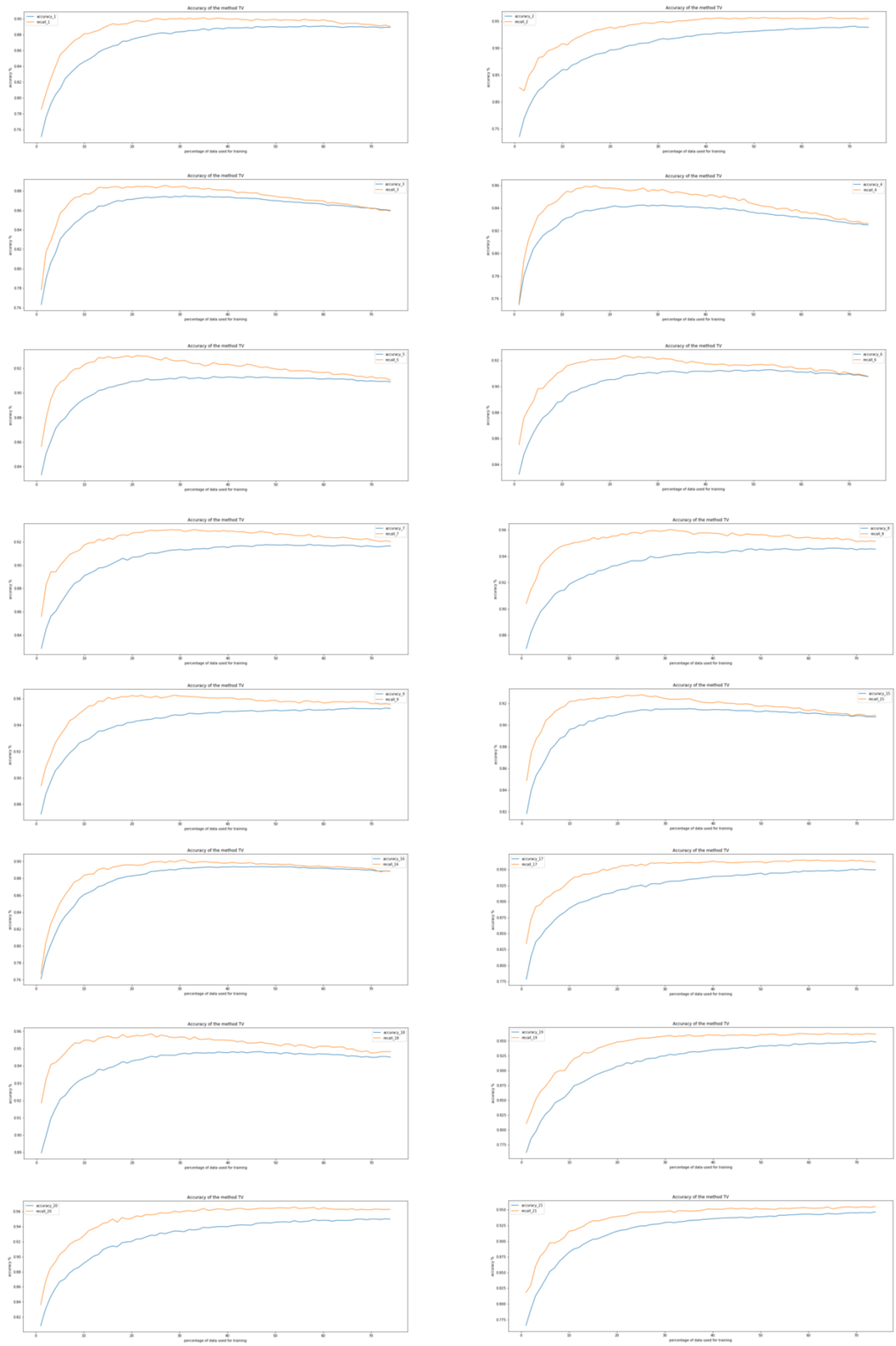
The results validated the idea that user behaviour can be identified, and a machine learning approach was demonstrated that can be applied to extract useful behaviour data from electrical load measurements or binary datasets defining user behaviour over time. These methods can be used by home automation companies to be integrated as the basis of their user interface. Home automation installers can use it to offer greater personalisation to their clients while requiring less of their time. The findings of the report could be used in the process of smart home planning by designers installing a few sensors in their client's homes for 1-3 months' time and based on the data analytics design the system with their behaviour in mind. This is a short enough time in case the customer is planning a full house renovation and tailoring the works and designs based on their behaviour can be a strong additional selling point.

Over all the goals of the project were mainly met, apart from creating a working mock-up of the user interface that can be tested with the available dataset due to the limitations in time and coding expertise. The concept validation and research were successful and yielded actionable results.

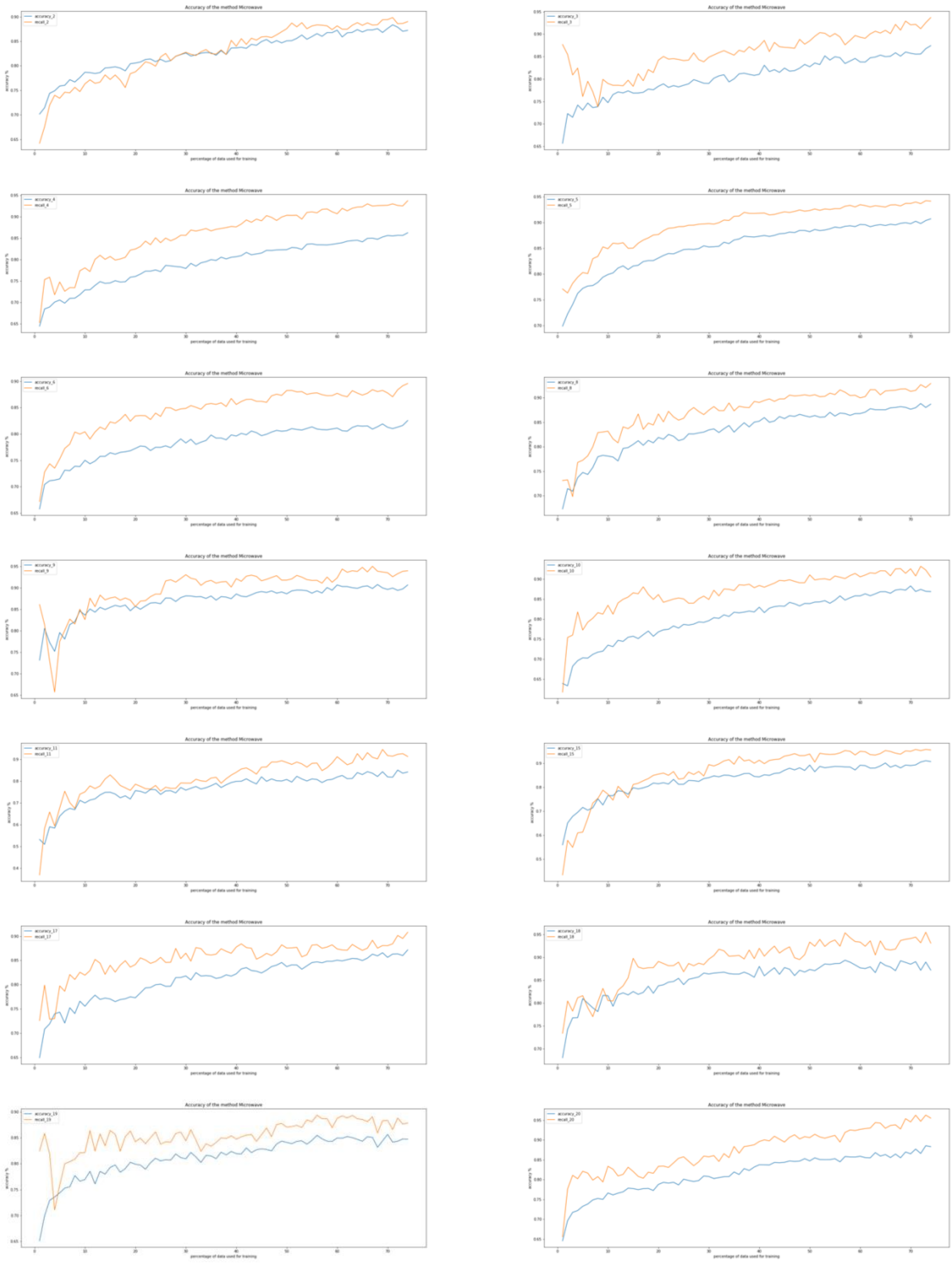
Appendices

Find the main python codes used and further figures here:
https://drive.google.com/drive/folders/1n5AiCfFt_rPtFtyWvd_xSjVWUtyed3QM?usp=sharing

1. Figure 6 Breakdown for individual houses



2. Figure 8 breakdown for individual houses



3. Visualisation of predictions

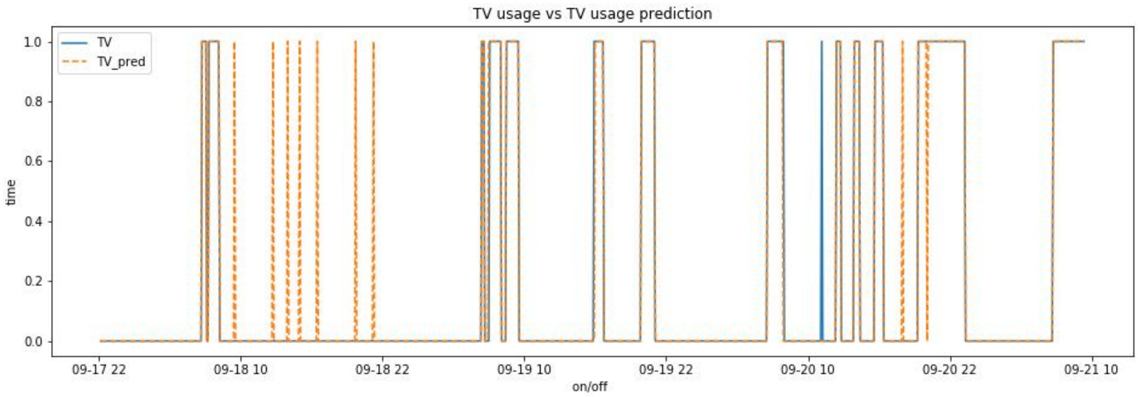
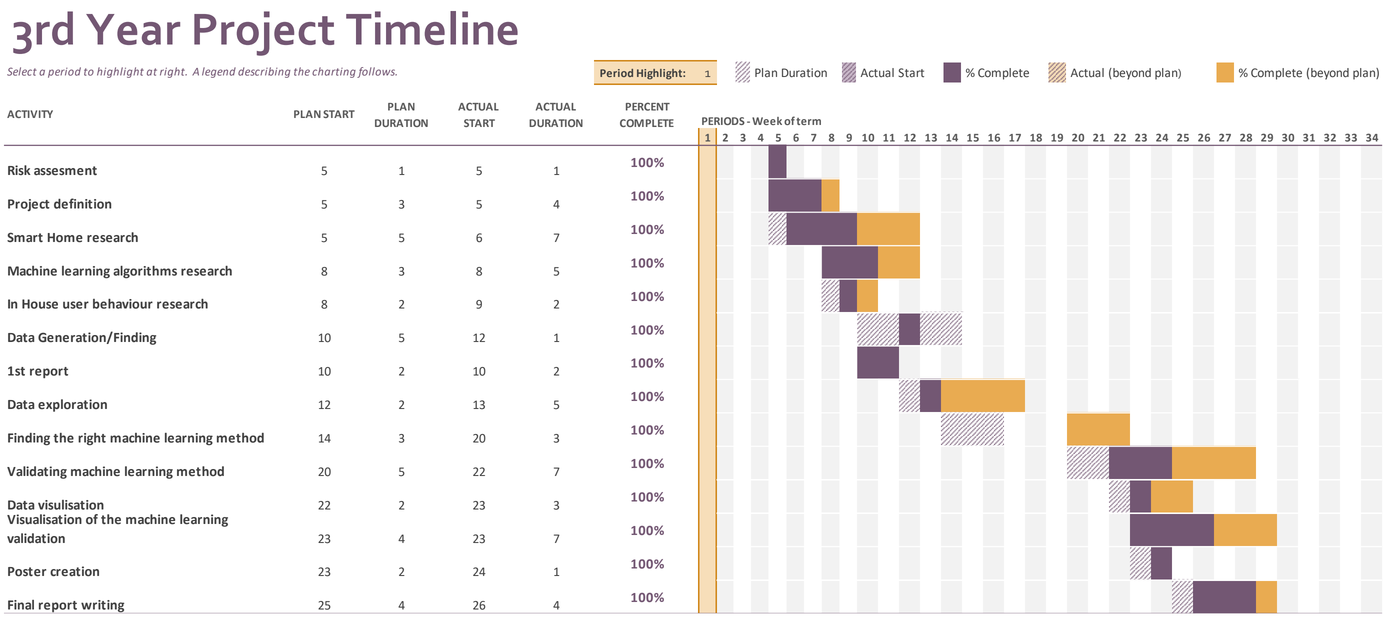


Figure 11 House 5; 5 minute interval resampling of the load measurement data

4. Timeline for the Project



ACTIVITY	PLAN START	PLAN DURATION	ACTUAL START	ACTUAL DURATION	PERCENT COMPLETE
Risk assesment	5	1	5	1	100%
Project definition	5	3	5	4	100%
Smart Home research	5	5	6	7	100%
Machine learning algorithms research	8	3	8	5	100%
In House user behaviour research	8	2	9	2	100%
Data Generation/Finding	10	5	12	1	100%
1st report	10	2	10	2	100%
Data exploration	12	2	13	5	100%
Finding the right machine learning method	14	3	20	3	100%
Validating machine learning method	20	5	22	7	100%
Data visulisation	22	2	23	3	100%
Visualisation of the machine learning validation	23	4	23	7	100%
Poster creation	23	2	24	1	100%
Final report writing	25	4	26	4	100%

5. Dataset Details

Acknowledgements

This work has been carried out as part of the REFIT project (Personalised Retrofit Decision Support Tools for UK Homes using Smart Home Technology, Grant Reference EP/K002457/1).

REFIT is a consortium of three universities - Loughborough, Strathclyde and East Anglia - and ten industry stakeholders funded by the Engineering and Physical Sciences Research Council (EPSRC) under the Transforming Energy Demand in Buildings through Digital Innovation (BuildTEDDI) funding programme. [1]

For more information see: www.epsrc.ac.uk and www.refitsmarthomes.org

REFIT: Electrical Load Measurements [20]

This Dataset ties into the previous one and details the electrical consumption of appliances monitored in each home.

THE FOLLOWING DATASET HAS BEEN CLEANED.

This has involved correcting the time for UK daylight savings, moving sections of Individual Appliance Monitor (IAM) columns to correctly match the appliance they were recording when a reset or householder moved them, NaN values have been forward filled (Please see the RAW dataset for non-forward filled values) and spikes of greater than 4000 Watts have been removed from the IAM values and replaced with zeros. There is also an additional issues column which is set to 1 if the sum of the sub-metering (IAMs) is greater than that of the household aggregate. In these cases, the data should be discarded or noted that there is a discrepancy. [20]

Exploring the Data

The refit datasets structure is built the following way: an xml files contains all the information about which house has what within it, and sensor types, start and end of measurements and rooms. The data points collected are in a separate csv file that links to the xml file through an id. This dataset was good insight into how the houses were set up.

It was also useful to show correlation between inside and outside temperature and gas consumption, inside and outside light levels, revealing the times lights were turned on in the house.

The downside of this dataset was that all readings were taken at a 30-minute interval. To assess the behaviour of the home occupants the sampling frequency of the sensors should be higher.

This is where the electrical load measurements [20], became useful as this had a sampling frequency of 5 seconds on average. This dataset can be easily resampled to a desired rate.

6. Random forest Regression

The Random Forest Tree regression can be used for the predicting actual settings, like temperature levels (Figure 12). Music volume or other settings for which different levels have to set on a linear scale.

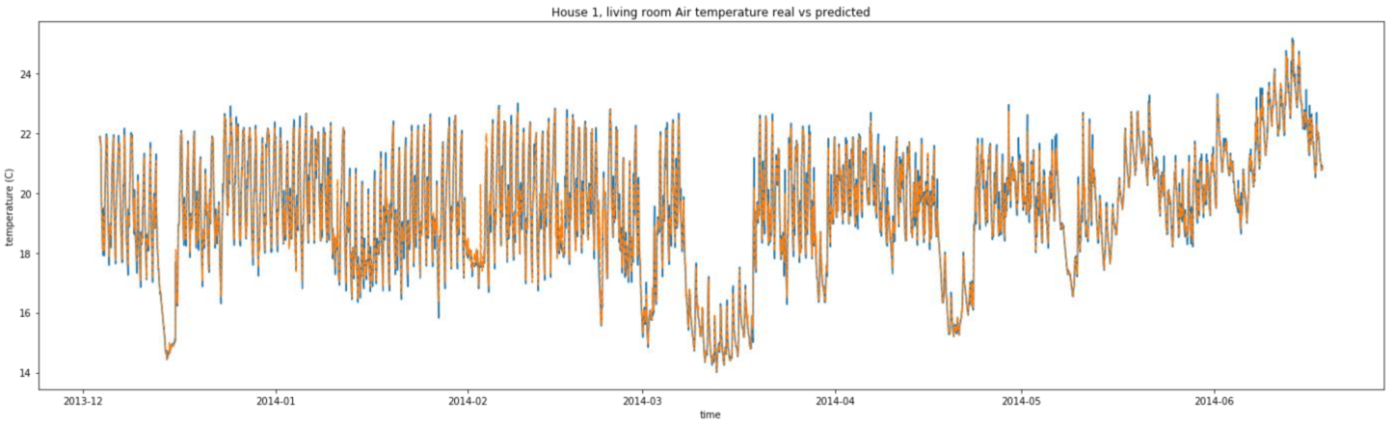


Figure 12 RFT regression applied to living room temperature from House 1

Bibliography

[1] S. Firth, T. Kane, V. Dimitriou, T. Hassan, F. Fouchal, M. Coleman and L. Webb, "REFIT Smart Home dataset," [Online]. Available: https://dx.doi.org/10.17028/rd.lboro.2070091.

[2] Google, "Google Connected Home," Google, 2019. [Online]. Available: https://store.google.com/category/connected_home.

[3] Amazon, "Alexa smart home devices," Amazon, 2019. [Online]. Available: https://www.amazon.com/smart-home-devices/b?ie=UTF8&node=6563140011.

[4] Apple, "IOS - Home," Apple, 2019. [Online]. Available: https://www.apple.com/uk/ios/home/.

[5] Promelek S.A., “Life Evolution,” 2019. [Online]. Available: https://www.lifeevolution.ro.

[6] Elektra S.A, “BUILDING AUTOMATION Intelligent control and management of spaces designed for comfort, security and energy efficiency,” 2017. [Online]. Available: https://www.grupoelektra.es/en/soluciones/automatizacion-de-edificios.

[7] A. B. Brush, . B. Lee and R. Mahajan, “Home automation in the wild: challenges and opportunities,” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011.

[8] [Online]. Available: https://www.zigbee.org.

[9] S. B. Aru, “Wireless Home Automation System Using Zigbee,” *International Journal of scientific & Engineering Research,* vol. 5, no. 12, pp. 133-138, 2014.

[10] J. Tidd, “Development of Novel Products Through Intraorganizational and Interorganizational Networks,” *Journal of Product Innovation Management,* vol. 12, no. 4, pp. 307-322, 2010.

[11] KNX.org, “KNX introduction,” 2018. [Online]. Available: https://www.knx.org/knx-en/for-professionals/What-is-KNX/A-brief-introduction/index.php. [Accessed 2018].

[12] KNX.org, “KNX wireless products,” [Online]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=12&ved=2ahUKEwj0s_HDgrveAhWJmLQKHSyIAf4QFjALegQIARAC&url=https%3A%2F%2Fwww.knx.org%2FwGlobal%2FwGlobal%2Fscripts%2FaccessDocument.php%3FforceDownload%3D0%26document%3D%252FwAssets%252Fdocs%2.

[13] A. B. Gökbayrak, S. Kilivan, S. Akin, A. Çelebi and O. Urhan, “Wireless sensor network-based extension to KNX home automation system,” *Turkish Journal of Electrical Engineering and Computer Sciences,* vol. 24, no. 5, pp. 3652-3663, 2016.

[14] Thinka, “Thinka Feature overview,” 2019. [Online]. Available: https://thinka.eu/en/product-info.html.

[15] Pairot, “Product description,” 2019. [Online]. Available: https://www.xxter.com/pairot/en/.

[16] T. Hargreaves and C. Wilson, “Who uses smart home technologies? Representations of users by the smart home industry,” *ECEEE Summer Study Proceedings,* Vols. Rethink, Renew, Restart , pp. 1769-1780, 2013.

[17] M. Rao, “Big Tech in the Smart Home,” *CB Insights,* 2018.

[18] Thinka, “Thinka for KNX description,” 2018. [Online]. Available: https://thinka.eu/en/product-info.html.

[19] Control4, “Whole Home Solutions,” Control4, 2019. [Online]. Available: https://www.control4.com/solutions/whole-home.

[20] D. Murray, J. Liao, L. Stankovic, V. Stankovic, R. Hauxwell-Baldwin, C. Wilson, M. Coleman, T. Kane and S. Firth, “A data management platform for personalised real-time energy feedback,” in *Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting*, 2015.

[21] F. Issi and O. Kaplan, “The determination of load profiles and power consumptions of home appliances,” *Energies,* vol. 11, no. 3, 2018.

[22] “sklearn.linear_model.LogisticRegression,” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.

[23] "1.6. Nearest Neighbors," [Online]. Available: https://scikit-learn.org/stable/modules/neighbors.html.

[24] "sklearn.ensemble.BaggingClassifier," [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html.

[25] "sklearn.ensemble.RandomForestClassifier," [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Ducheasnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825-2830, 2011.

[27] A. Brenon, F. Portet and M. Vacher, "ARCADES: A deep model for adaptive decision making in voice controlled smart-home," *Pervasive and Mobile Computing,* vol. 49, pp. 92-110, 2018.

[28] C. Withanage , R. Ashok , C. Yuen and K. Otto, "A comparison of the popular home automation technologies," *2014 IEEE Innovative Smart Grid Technologies - Asia, ISGT ASIA 2014,* no. May 2015, pp. 600-605, 2014.