The Bartlett School of Environment, Energy and Resources

MSc ESDA Title Page

UCL Candidate Code: The Turtles (QFDF0, RQHB8, TYGR3, TWCD3)

Module Code: BENV0091

Group Number: 4

Module Title: Energy Data Analysis

Coursework Title: Investigating the Effect of Temperature Variables on Households Classification: A Case Study of London Smart Meter Data

Module Leader: Patrick de Mars

Date: 13 December 2021

Word Count: 4718

By submitting this document, you are agreeing to the Statement of Authorship:

We certify that the attached coursework exercise has been completed by us and that each and every quotation, diagram or other piece of exposition which is copied from or based upon the work of other has its source clearly acknowledged in the text at the place where it appears.

We certify that all field work and/or laboratory work has been carried out by us with no more assistance from the members of the department than has been specified.

We certify that all additional assistance which we have received is indicated and referenced in the report.

Please note that penalties will be applied to coursework which is submitted late, or which exceeds the maximum word count. Information about penalties can be found in your Course Handbook which is available on Moodle: https://moodle.ucl.ac.uk/mod/book/view.php?id=2234010

- ***Penalties for late submission***
- ***Penalties for going over the word count***

In the case of coursework that is submitted late and is also over length, then the greater of the two penalties shall apply. This includes research projects, dissertations and final reports.

Table of Contents

List of Figures

List of Tables

1. Introduction – Overview and Objectives

The UK's ambition to achieve Net Zero by 2050 will see substantial changes and transformations within the energy system. These changes will bring new challenges and tough policy decisions for the UK Government. The Climate Change Committee (CCC) has estimated that Net Zero will cost the UK £1.4tn by 2050. HM Treasury has also warned that new taxes will be introduced as part of the Net Zero transition due to the transition from fossil fuels to low carbon technologies decreasing fuel duty and vehicle excise duty which raised £37bn in 2020 (Financial Times, 2021).

With this change, the issue of household fuel poverty will become a more pertinent issue with the concerns that the poorest households will be left behind in the net-zero transition. The Department for Business, Energy and Industrial Strategy (BEIS) defines a household in fuel poverty if they are on a lower income and unable to heat their home for a reasonable cost (BEIS, 2021). The proportion of households in fuel poverty in 2019 was 13.4% which is over three million households in the UK.

Understanding household behaviour will also be crucial towards net-zero as this is considered an effective way to improve energy efficiency and promote energy conservation (Zhou and Yang, 2019). The challenges, predicting and forecasting household consumptions based on socioeconomic groups will be essential to comprehend how climate change and net-zero will affect their consumption behaviour. With significant developments in big data and data mining, researchers have used unsupervised data clustering and frequent pattern mining analysis on energy-time analysis and Bayesian network prediction for energy usage forecasting (Singh and Yassine, 2018). The paper presented an intelligent mining model to compute and visualise energy time series data to understand several energy household consumption trends.

Furthermore, there is a growing body of literature on various methods of predicting the household groups based on energy consumption, behaviour, socioeconomic group, and area of residency. This paper aims to predict household groups based on electricity consumption using several classification algorithms and comparing their effectiveness.

Our research will also evaluate whether incorporating temperature within the model will better predict groups of affluent and non-affluent household energy consumers using electricity consumption data. In theory, applying seasonal temperature to household energy consumption should increase accuracy in classifying the households. We discuss the suitability of the models and we examined and their effectiveness in their accuracy. An interactive display, providing an overview of our study, data exploration and model performance evaluation, is available online at:

https://theturtles.shinyapps.io/eda-group-project

2. Literature – Literature Review

Research into household energy consumption has always been a pertinent and essential topic in planning future energy systems and understanding system constraints within the energy system at a specified time. The studies typically fall into two categories, households classification and forecasting energy consumption.

There has been a growing study about classifying households based on electricity consumption. Nielsen and Nørgård (2009) argued that household classification could help the government manage electricity taxes and stated that household size might affect electricity consumption. Amiruddin et al. (2020) compared some machine learning models to group households using energy consumption data. They suggested that Logistic Regression, K-Nearest Neighbour (KNN), Support Vector Machines (SVM), and Decision Tree models are recommended to classify households based on their accuracies.

As early as 2012, researchers have been analysing energy usage data and behavioural patterns to improve accuracy in demand forecasting models. Campillo et al. (2012) processed energy usage data, behavioural patterns from 5000 end-users in different Swedish DSO areas and physical conditions for the facilities. The paper noted that the model results are valid for testing different policies and flexible demand strategies using trend, similar day, end-use and econometric modelling.

Bonetto and Rossi (2017) analysed several approaches in solving the problem of power demand forecasting in residential microgrids. They concluded that when using the Auto-Regressive Moving Average (ARMA) model, Support Vector Machines (SVM), Long Short-Term Memory (LSTM) and Nonlinear Auto-Regressive (NAR) all outperformed ARMA. They also noted that adopting a hybrid approach with the use of NAR for short term horizons and SVM for long term forecasting intervals was the optimum solution.

In a particular instance, forecasting single-household residential energy consumption was also predicted using Support Vector Regressions (SVR) modelling for daily and hourly data granularity (Zhang et al., 2018). Jin et al. (2022) noted that SVR is limited as the model cannot be constructed until the whole dataset is available. Therefore, depending on data availability might not be effective compared to other models. Deep residual neural networks have also been applied in forecasting day-ahead household electrical energy consumption. Multiple test cases show that the proposed model provides accurate load forecasting results (Kiprijanovska et al., 2020). Yan et al. (2018) used a hybrid deep learning neural network framework that combines a convolutional neural network (CNN) with LSTM to predict short-term power consumption for individual households.

One paper used additional variables to forecast electricity consumption in developed and developing countries (Ardakani and Ardehali, 2013) found that using socioeconomic indicators led to more accurate forecasting in electrical energy consumption for both developing and developed countries in their selection. Optimised regression and artificial neural network (ANN) models were developed to forecast EEC between 2010-2030. ANN models are known to be a robust methodology for

modelling hourly and daily consumption and load forecasting. Rodrigues, Cardeira and Calado, (2014) defined an ANN architecture and a training algorithm to create a robust model to forecast energy consumption in a typical household. Tracking 93 households' daily consumption between February 2000 and July 2001 found using an ANN approach provided a reliable model for forecasting electric energy consumption.

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance where the popularity of the algorithm has increased significantly within the scientific community. Hadri et al. (2021) investigated XGBoost, LSTM and SARIMA. It noted the challenging trade-offs between embedded forecasting model training and processing for being deployed in smart meters for electricity consumption forecasting. They also concluded that XGBoost outperforms univariate and multistep forecasting well against ARIMA variants, which showed better performance for multivariate scenarios.

3. Methodology

Research Workflow

Figure 1 shows how we conducted our research. Initially, smart meter, household information, and temperature datasets were merged to create a single dataset containing all crucial variables. Then, we analysed our dataset to find general ideas, such as how the electricity consumption differs for each household group and how the temperature fluctuates. Some of the visualisation results from this step are displayed interactively on our website.

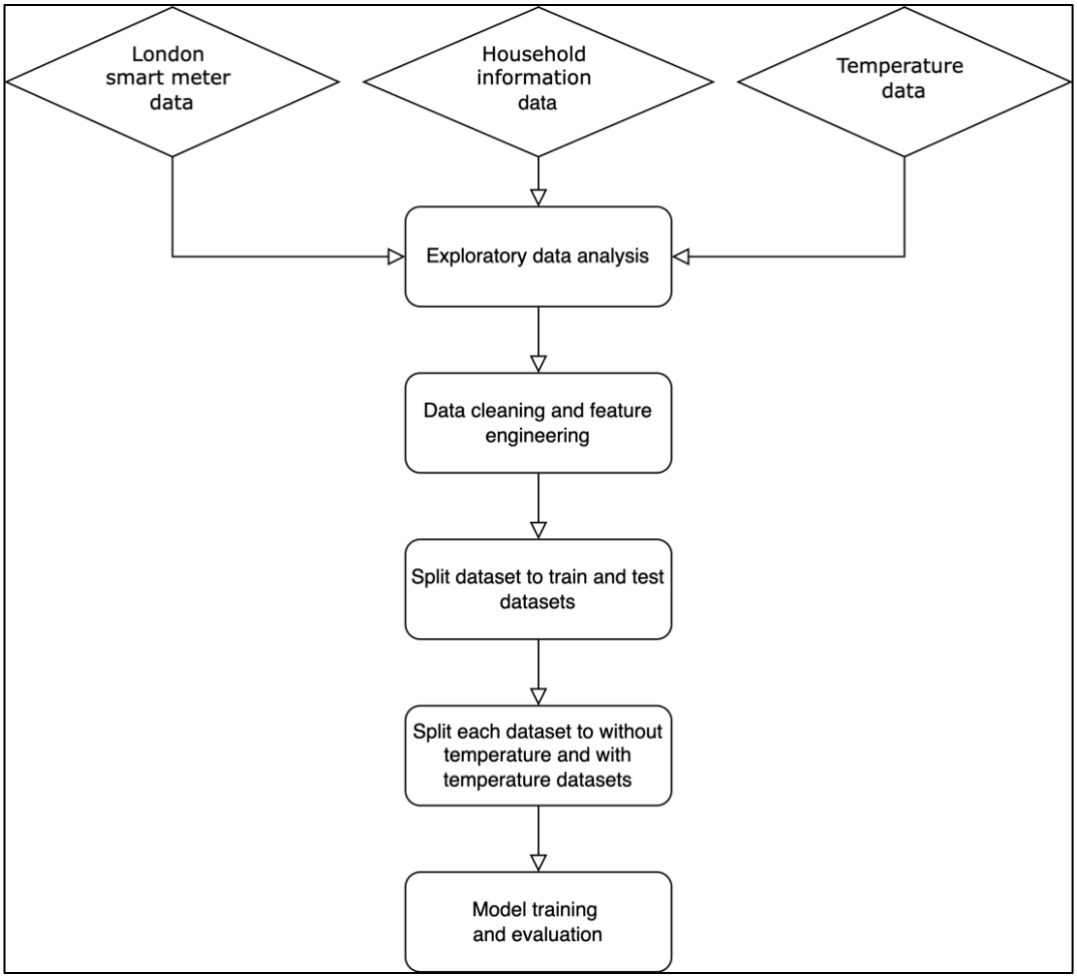

Figure 1. Research Workflow.

As we aim to develop machine learning models to classify household groups and compare the effects of incorporating temperature into the dataset, we cleaned the data and created new features by dividing the daily electricity consumption with daily temperature data. These new variables are expected to represent the effect of temperature variation. These processing steps widened our dataset since 53 new features represented each week's data. After that, the dataset was split into train and test datasets before training the model. Three machine learning algorithms were used – logistic regression, random forest, and extreme gradient boosting. Finally, we investigated our model performance to answer our research formulation about temperature effects.

Data Description

Three data sets were chosen for this project and data analytics: Low carbon London, Acorn low carbon London, and Heathrow weather data. The low carbon London data set was collected over three years from 2011 to 2014 with 4500 participants on a fixed price per while 1,100 were on a dynamic time, tariff as stated in Research report for Citizens Advice. The Acorn group was an addition to the low carbon London data set. Within this data set, each household was put into one of seventeen groups. These are then sorted into five categories. Each of these is then sorted into one of the three groups used for the data set being Affluent, Comfortable and Aversity. The Heathrow weather data set is a continuous daily collection of weather information.

The low carbon London data set was in the following format, Household ID, Type of tariff either Standard or Dynamic Time of Use, Date and time and KWH consumed per hour. New empty data frames were created and extracted 'date and time' and 'KWH per half hour' in their raw form. NA's were then removed, and zero daily consumption values were also removed. This was done due to the nature of the overall question and the linked data used. This will be further explained below. It also allowed a more straightforward interpretation and better-captured variations within the data.

The second data set used was the Acorns Low carbon data set. This had the following format, Household ID, type of tariff either Standard or Dynamic Time of Use, and the Acorn group that each household belongs to the following Affluent, Comfortable and Adverse. The final data set was a weather data set collected from Heathrow. The following data was used from this Date, Max, minimum and average temperature. With all the data set acquired, they were then joined. The low carbon London and the acorn group was joined via the Household ID. The combined data set was then joined to the temperature data set using the date.

The final data set was left with Household ID, Date, Acorn Group, Daily KWH usage, Average Daily temperature. This was then expanded with the following columns using lubridate, days, weeks, months and years for future visualisations and grouping.

Subsequently, consumption values of each half-hour were averaged throughout each day and used for the classification algorithm. Using averages improves the robustness of the model because households have different starting points. Meanwhile, for visualisation purposes, the consumption values of each half-hour were summed to get the daily total, as this provides a more interpretable visualisation.

For the classification algorithm, daily averages were used to normalise the data for each household over a 24-hour period. This was due to households having different starting dates for the study. The use of averages increased the robustness of the model, allowing a more straightforward interpretation and better-capturing variations within the data.

A different method was used for the visualisation of the data. A sum of the half-hourly values was used to get a daily total. This provided a more precise visualisation of daily trends.

Classification Methodology

Logistic Regression

Logistic regression is a type of classification algorithm and solves the probability of success and failure. The algorithm learns a linear relationship from the datasets and creates non-linearity characteristics through Sigmoid functions. Logistic regression models were fit to data obtained under experimental conditions. Research have used logistic regression to forecast energy demand. For example, Shaikh and Ji (2016) forecasted natural demand in China using logistic modelling and applied a Levenberg-Marquardt algorithm to estimate the parameters of the logistic model.

Pregibon, (1981) noted that the use of logistic regressions methods includes the analysis data obtained in observational studies. This is evident our dataset as the data used was electricity household consumption from smart meters.

Applying the maximum likelihood function is the common method of fitting logistic regression models but, depending on the quality of the dataset, is susceptible to being highly sensitive.

Logistic regression can be expressed as:

$$\log\left(\frac{p}{1-p}\right) = y$$

Where $p/(1-p)$ is an odd ratio and y is the linear model, and p is the probability of success. The value of p is between 0 and 1.

Random Forest

Random forest is an ensemble of decision trees where it builds and combines multiple decision trees increasing accuracy in predictions. Predictions are calculated by summing the predictions of the ensemble, Svetnik et al. (2003) investigated the effectiveness of random forest and concluded that the tool is an extremely powerful performance that is among the most accurate methods to date.

Random Forest has previously been used to predict electricity usage. Wang et al., (2018) applied Random Forest to predict the hourly electricity usage of two educational buildings in North Central Florida. The models' performance was evaluated with the models being trained with different parameter settings. The paper also found that the Random Forest model being trained with yearly and monthly data forecasts could be improved by considering their energy behaviour changes during different semesters.

Boosting and XGBoost

Generally, tree boosting is a commonly used machine learning method, XGBoost is a scalable end-to-end tree boosting system. Forecasting day-ahead wind power has always been a big challenge due to the complexity of wind behaviour. Jiang et al. (2017) proposed a boosting algorithm and found that this algorithm performance provided a more accurate forecast than a ARMA model.

Chen and Guestrin (2016) noted that XGBoost scales beyond multiple scenarios using far fewer resources than existing systems. The algorithm is essentially decision trees in sequence where weights can be tuned which are then inputted into a decision tree. XGBoost is an efficient and scalable implementation of a gradient boosting framework (Friedman, 2001). Researches forecasted crude oil prices where parameters which are the factors affecting crude oil price will be interpreted by using XGBoost (Gumus and Kiran, 2017).

Classification Metrics

There are some available classification metrics for classification purposes, such as precision, recall, F1 score, and accuracy (Lever, Krzywinski and Altman, 2016). These indicators can be derived from the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 \times \left(\frac{precision \times recall}{precision + recall}\right)$$

Where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative. The following results are:

- True Positive is the number of predictions labelled as positive, and their actual values are positive.
- True Negative is the number of predictions labelled as negative, and their actual values are negative.
- False Positive is the number of predictions labelled as positive, but their actual values are negative.
- False Negative is the number of predictions labelled as negative, but their actual values are positive.

4. Results

Exploratory data analysis

Data transformation and visualisation have been performed to understand the data better.

Firstly, the total electricity consumption has been plotted with the average temperature for every day between 23 November 2011 and 29 February 2014, which constitutes the entire length of the consumption data (Fig. 2). Furthermore, each acorn group's total daily consumption and average temperature for the same period have been plotted separately to visualise other trends and discrepancies.

Secondly, the period of 23 November 2011 and 29 February 2014 has been averaged to produce two plots for each Acorn group: one of the average daily temperature and total consumption for one month, and another of the average weekly temperature and total temperature for one year.

Consumption and temperature plots: full length

As expected, we can observe a strong reverse correlation between consumption and temperature values (Fig. 2). This is especially visible in consumption peaks during very low temperatures in the winter season. Moreover, the data seems to be following a cyclical trend with similar highs and lows in both temperature and consumption associated with seasons – lower consumption and higher temperature in the summer and higher consumption and lower temperatures in winter.

We can see very high consumption values, with daily totals reaching over 100 kWh and, in some cases, over 300 kWh. Though, as it is indicated by the dashed line (Fig. 2) at the mean value of 10 kWh (Table 1), the majority of total daily values are much lower. Furthermore, consumption values over 20 kWh constitute only 9.4% of total observations.

Opacity in the plot indicates the density of values. As seen from the middle and bottom parts of Figure 2, most consumption values over 20 kWh are focused towards the cold season, while the values below 20kWh stay consistent throughout the whole timeframe. The general lack of observations causes Low-value density during the first half of 2012.

Acorn Group	Daily Mean (kWh)	Weekly Mean (kWh)
Affluent	11.49	88.96
Comfortable	9.98	77.21
Adversity	8.54	66.12
Overall	10	77.4

Table 1. Mean values of total daily and weekly consumption.

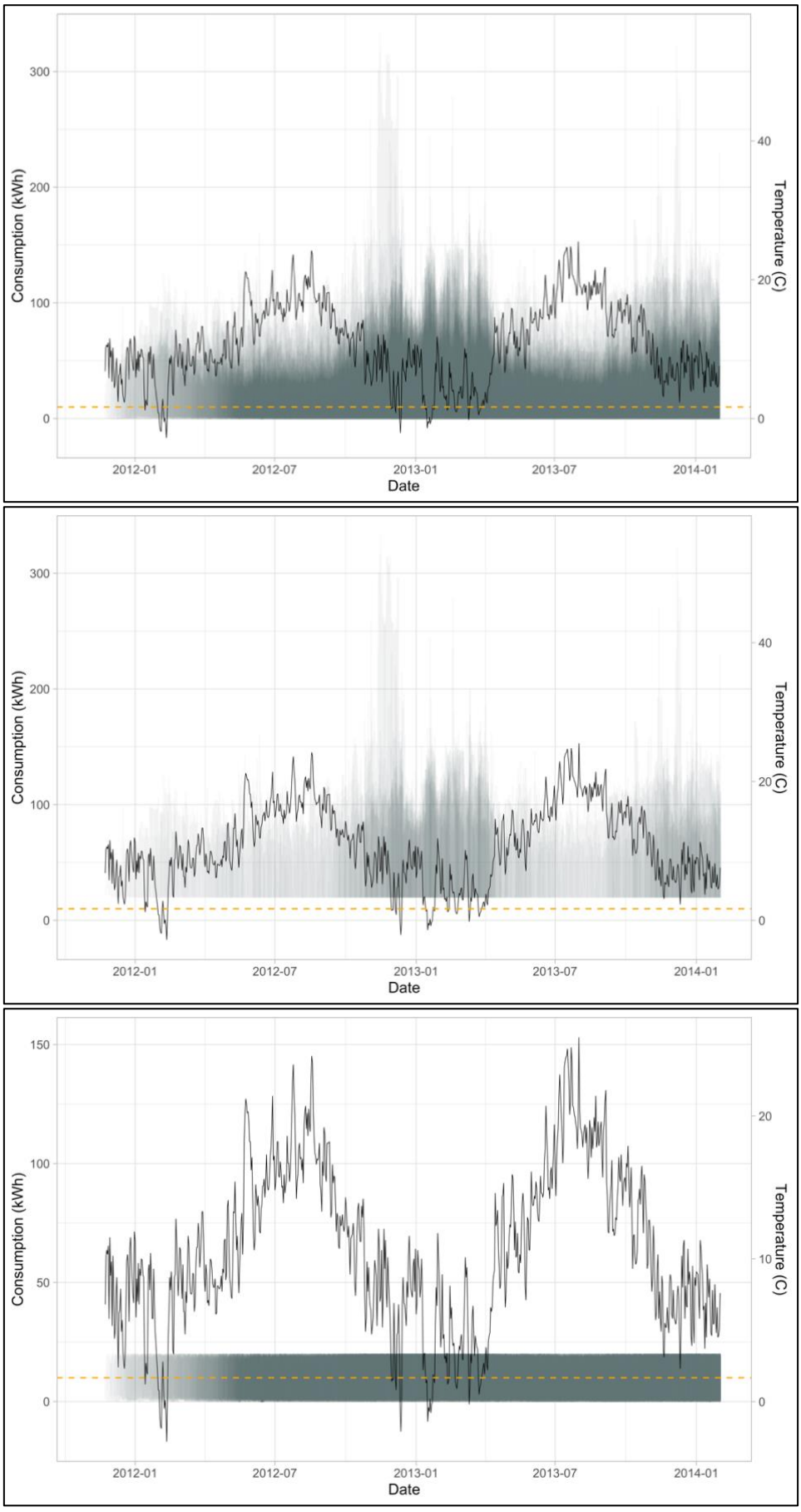

Figure 2. Plot of total electricity consumption with average temperature for every day between 23 November 2011 and 29 February 2014. Temperature values are shown by the black curve; orange dashed line is at the mean of all consumption values (10kWh).

Following these observations, an assumption can be made regarding the primary heating source of the households included in the dataset. According to Intertek (2012), the average annual electricity consumption was 3,638 kWh for households without electric heating and 5,431 kWh for households with primary electric heating. For a single day, that would equate to approximately 9.9 kWh and 14.8 kWh, respectively. Only 18.2% of overall observations are above 14.8 kWh.

Considering that the average value of electricity consumption for households without electric heating is close to the mean of the total daily values of the Low Carbon London dataset, it is safe to assume that the majority of the households in the dataset are not using electricity for primary heating.

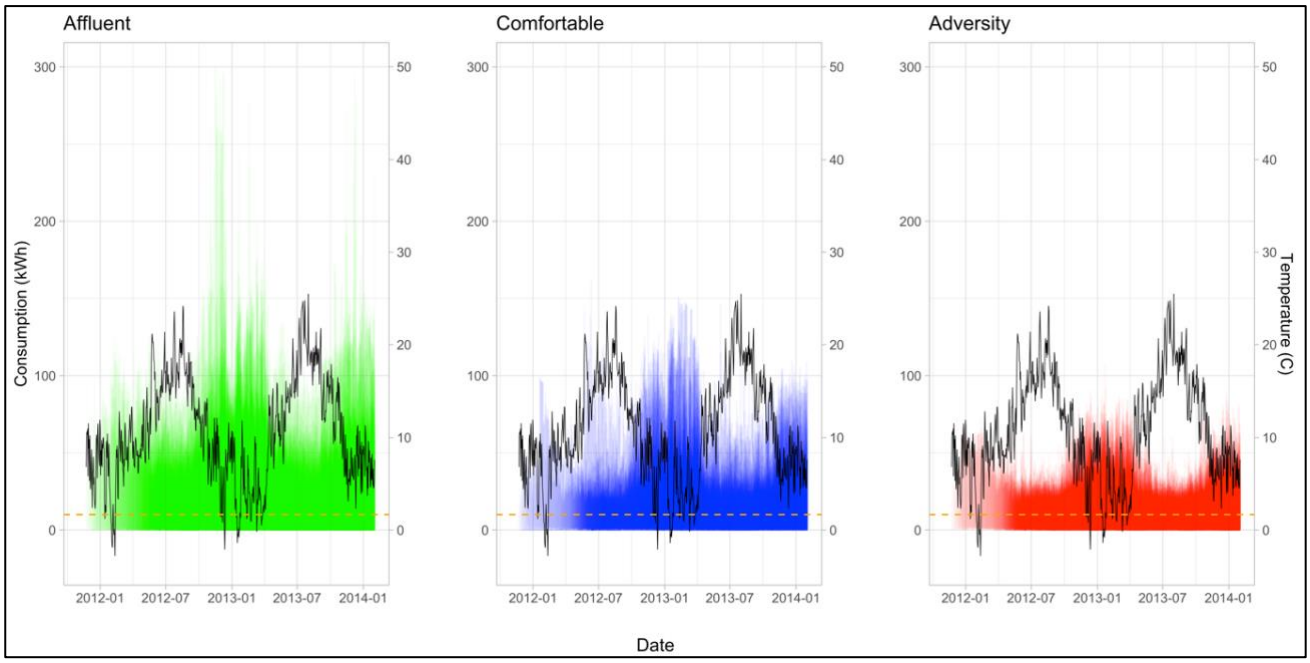

Figure 3. Plot of total electricity consumptions for each Acorn group with average temperature for every day between 23 November 2011 and 29 February 2014. Temperatures are shown by the black curve; orange dashed line is at the mean for consumption (10kWh).

Plots of consumption values for the three Acorn groups (Fig. 3) show a general increase in consumption from Adversity to Affluent. We can also see that all anomalous peak values over 150 kWh are related to the Affluent group. One of the possible causes for such anomalous peaks could be excessive lightning and cooking during Christmas.

Another important observation from Figure 3 is that Affluent households show an increase in consumption earlier in the cycle, which leads to consumption peaks thickening from Adversity to Affluent. This can be seen in October for each year when Affluent households have already reached their peak values, while Comfortable households are approaching their peak and Adversity households are still far from it.

Consumption and temperature plots: averaged

The previous observations are further supported by Figure 4, which shows the difference between each Acorn group's total daily consumptions averaged for one month. The average daily consumptions of the dataset are close to the UK average daily consumption for households without electric heating. Consumption values of the Affluent group are below 12 kWh and Adversity below 9 kWh. The Comfortable group shows values very close to the overall mean of the dataset.

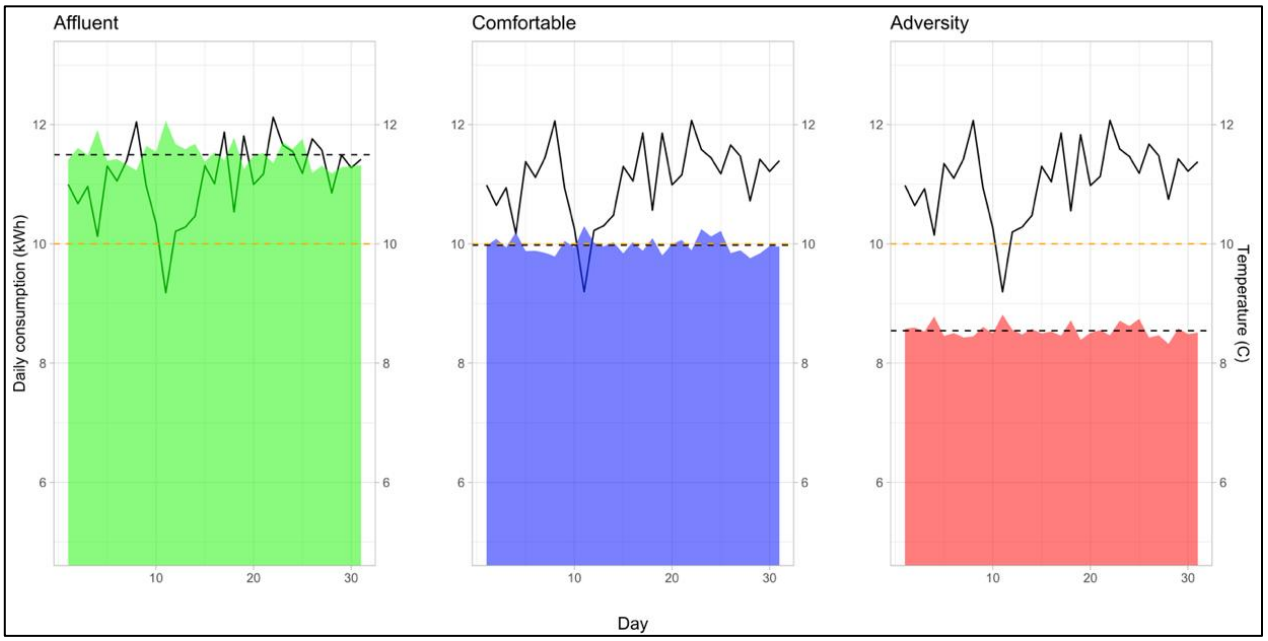

Figure 4. Daily total consumption of each Acorn group vs temperature, averaged for one month. Temperatures are shown by the black curve; orange dashed line is at the mean forl consumption; black dashed line is at the consumption mean specific to the Acorn group.

Weekly temperature and total consumption averaged for one year

An observed cyclical pattern (Fig. 2) in temperature and consumption allows us to assume that averaging between the years would avoid significant loss of information. Averaged daily totals were further summed to get total weekly consumption values, sacrificing the resolution in favour of the general trend.

Figure 5 shows each Acorn group's weekly consumption values that support previously made observations. The consumption levels strongly correlate with the temperature, with overall consumption and its sensitivity to temperature growing from the Adversity to the Affluent group.

Furthermore, we can also observe that the consumption peaks grow higher and wider as the household affluence increases. Following our assumption that the majority of households do not use electricity for heating and the strong correlation between the temperature and sunlight (Pudovkin, 2004), it can be concluded that the variation is mainly due to the change in lightning behaviour and some secondary heating.

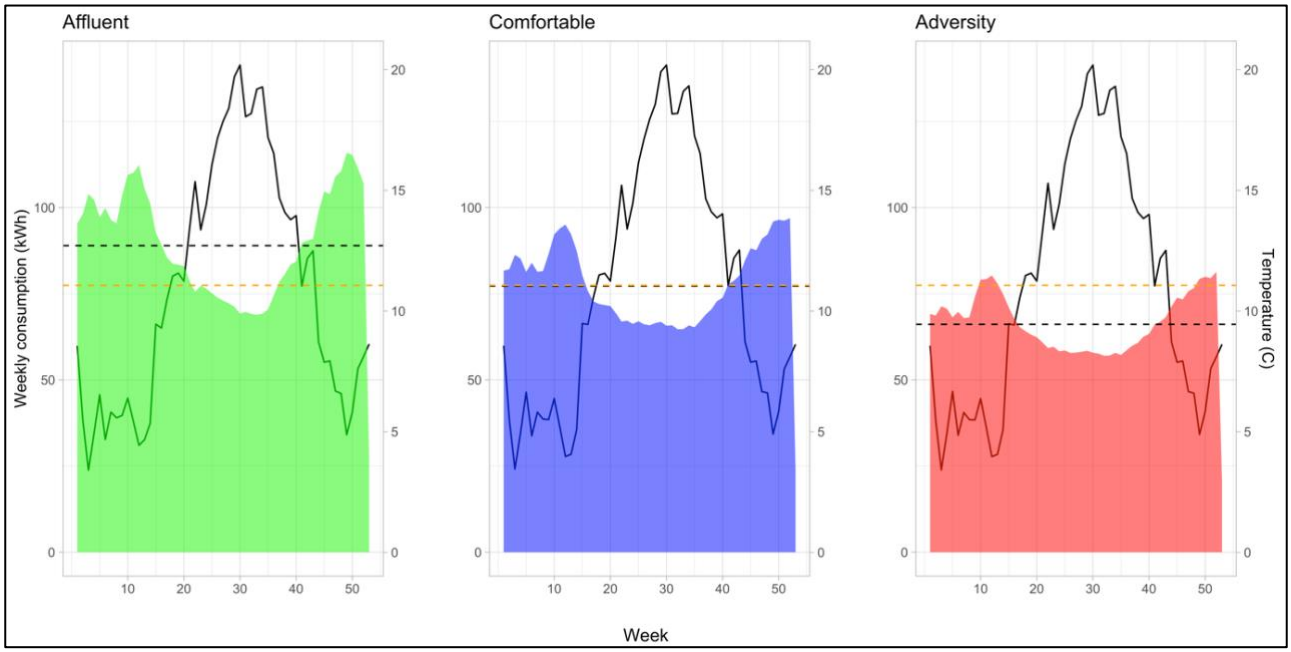

Figure 5. Weekly total consumption of each Acorn group vs temperature, averaged for one year. Temperatures are shown by the black curve; orange dashed line is at the consumption mean (75kWh); black dashed line is at the consumption mean for each Acorn group.

Data preparation and feature engineering

As mentioned in the Methodology section about how we pre-processed the data, we initially combined two datasets, electricity consumption and household information. Table 2 shows the samples from the joined dataset of electricity consumption and household information. So, we have the electricity consumption record for each household for each half-hour. This kind of dataset to classify households becomes a challenge in research since we want to predict the household's group based only on their electricity consumption variations.

Household ID	Datetime	Electricity consumption per half hour (kWh)	Group
MAC000036	2012-12-09 18:30:00	0.045	Affluent
MAC000036	2012-12-09 19:00:00	0.044	Affluent
MAC000036	2012-12-09 19:30:00	0.066	Affluent
MAC000036	2012-12-09 20:00:00	0.252	Affluent
MAC000036	2012-12-09 20:30:00	0.043	Affluent

Table 2. Samples of joined electricity consumption and household information dataset.

As the average temperature dataset is provided on a daily basis, we also need to summarise the electricity consumption in the same way so that we can join both of them. There are several possible grouping functions that we can use, such as sum, average, minimum, and maximum.

However, the problem with our dataset is that each household has a different starting and ending date. We have tried both the sum and average grouping methods to build the model. They ended up with similar results. However, the average summarising approach gives a more consistent result.

We decided to utilise the average method to build our models. Before calculating the average, some missing values were found, and they were dropped when averaging the values. Table 3 shows the samples from the joined dataset from electricity consumption, household information, and temperature.

Household ID	Date	Average electricity consumption per half hour (kWh)	Average temperature (°C)	Group
MAC000036	2011-12-07	0.07633333	7.4	Affluent
MAC000036	2011-12-08	0.05985417	8.7	Affluent
MAC000036	2011-12-09	0.05525000	5.1	Affluent
MAC000036	2011-12-10	0.06291667	2.4	Affluent
MAC000036	2011-12-11	0.05418750	6.2	Affluent

Table 3. Samples from all datasets combined.

The next step is feature engineering. This step is essential in our research as we want to know the effects of adding temperature to the dataset to build the model. We created a new feature, electricity consumption per temperature, that can capture electricity consumption and temperature information by dividing the average electricity consumption per half hour with average temperature:

$$Electricity\ consumption\ per\ temperature = \frac{Average\ electricity\ consumption\ (kWh)}{Average\ temperature\ (°C)}$$

This feature can be grouped by daily, weekly, monthly, or other periods as long as the observation period of electricity consumption matches with the period of temperature observation. We will observe whether using this new feature helps to improve the model's performance. Table 4 shows the calculation result of this new feature from the samples in Table 3. We will build models from Table 4 and then compare them with the models built using electricity consumption data in Table 3.

Household ID	Date	Electricity consumption to temperature ratio	Group
MAC000036	2011-12-07	0.07633333	Affluent
MAC000036	2011-12-08	0.05985417	Affluent
MAC000036	2011-12-09	0.05525000	Affluent
MAC000036	2011-12-10	0.06291667	Affluent
MAC000036	2011-12-11	0.05418750	Affluent

Table 4. Samples of electricity to consumption calculation result.

The boxplots in Figure 6 shows interesting facts about our new feature. If we compare it with the electricity consumption variable, the new variable has values below zero. This is because the effect of temperature data can be lower than zero.

The other intriguing information is that the maximum value of the boxplot always increases from left (Adversity) to the right (Affluent). Based on the analysis from the given dataset, the Affluent group tends to consume more energy than the other two groups.

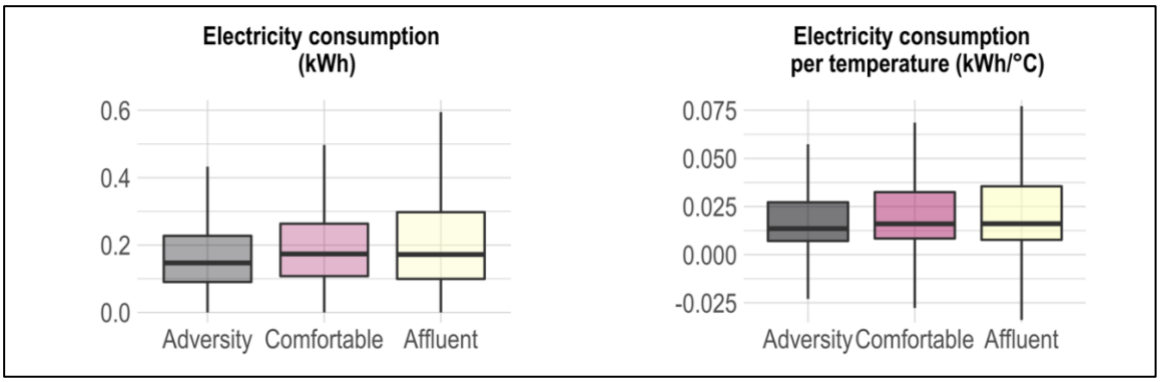

Figure 6. Distribution comparison between electricity consumption and electricity consumption per temperature features.

As an input to our models, we grouped our dataset on a weekly basis so that each household has 53 features from week one to week 53. The first week starts from the first day of the year. We have also tried using the standard ISO week that uses only 52 weeks, but the result was similar and but less consistent in accuracy.

We also utilised random seed for reproducibility. The training dataset has 4008 households for the train-test split, while the test dataset includes 1335 households. The ratio between train and test dataset is about 70:30. This condition holds for dataset that uses only electricity consumption and the one that uses electricity per consumption feature. We then have train and datasets ready to be used for model training.

Model training and evaluation

For each model, we will compare the results for each dataset that uses the electricity consumption feature (without temperature) and that uses consumption per temperature feature (with temperature).

Logistic Regression

We chose logistic regression as our baseline model. Figure 7 compares the confusion matrix of our logistic regression models. There is a slight improvement from the model that incorporates temperature features. The accuracy is improved by 3.6%, from 0.43 to 0.446. Furthermore, Tables 5 and 6 describe how the precision, recall, and F1 score improved by adding this temperature variable, except the comfortable group's precision. This might be because of the nature of the Comfortable group. The Comfortable group lies somewhere between the Adversity and Affluent classes, where the boundary is unclear. This might lead to our classification model's inconsistency.

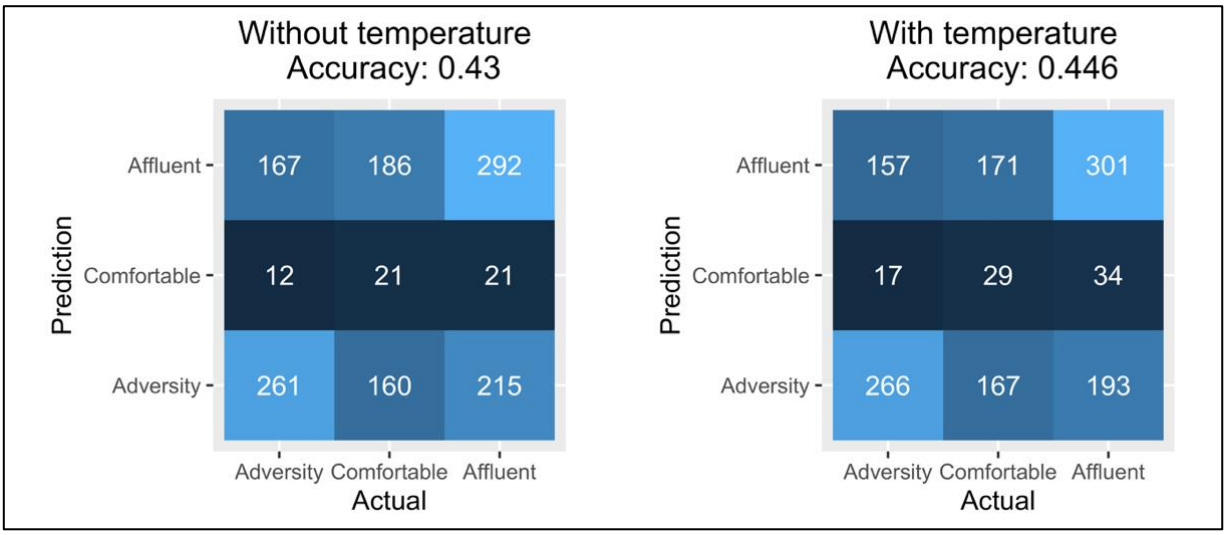

Figure 7. Logistic regression model's confusion matrices on test dataset with three classes.

Metrics	Group		
	Adversity	**Comfortable**	**Affluent**
Precision	0.4104	0.38889	0.4527
Recall	0.5932	0.05722	0.5530
F1 score	0.4851	0.09976	0.4979
Balanced accuracy	0.5871	0.55156	0.5578
Overall accuracy	0.43		

Table 5. Logistic regression model's evaluation metrics on test dataset (without temperature feature) with three classes.

Metrics	Group		
	Adversity	**Comfortable**	**Affluent**
Precision	0.4249	0.36250	0.4785
Recall	0.6045	0.07902	0.5701
F1 score	0.4991	0.12975	0.5203
Balanced accuracy	0.6012	0.51317	0.5818
Overall accuracy	0.4464		

Table 6. Logistic regression model's evaluation metrics on test dataset (with temperature feature) with three classes.

We conducted the same modelling approach but only two classes, Adversity and Affluent, with that uncertainty. The comfortable group's observations were temporarily dropped to measure how well the model can separate the other two classes. As shown in Figure 8 and Table 7, the overall performance increases when the number of classes is reduced to two. It still holds that the model that uses temperature variables provides better accuracy.

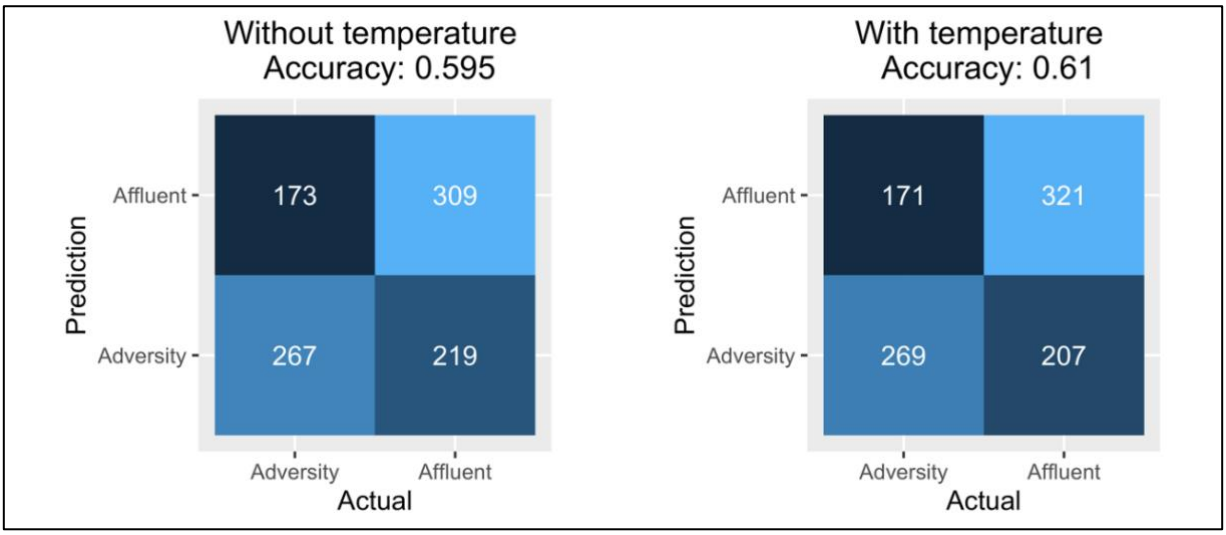

Figure 8. Logistic regression model's confusion matrices on test dataset with two classes.

Metrics	Without temperature	With temperature
Precision	0.5494	0.5651
Recall	0.6068	0.6114
F1 score	0.5767	0.5873
Accuracy	0.595	0.6095

Table 7. Logistic regression model's evaluation metrics on test dataset with two classes.

Random Forest

In terms of accuracy, random forest emerges as the top model in our study. Figure 9, Table 8, and Table 9 provide some intriguing results. For both without and with temperature models, the Comfortable group's recalls are higher than the previous logistic regression's metrics while its precisions decrease. For the comfortable group, the performance metrics that joins precision and recall, F1 scores, are improved from the logistic regression model. However, the adversity group's performance metrics decrease from the previous model. Based on these findings, our random forest model is more robust to classify the biased comfortable group than our logistic regression model.

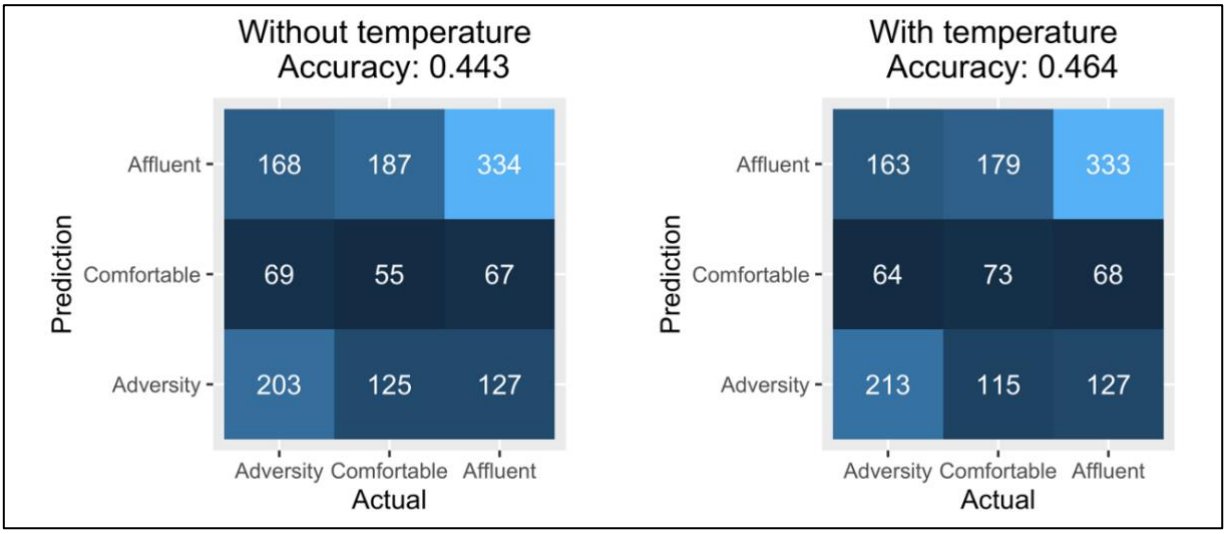

Figure 9. Random forest model's confusion matrices on the test dataset with three classes

Metrics	Group		
	Adversity	Comfortable	Affluent
Precision	0.4462	0.2880	0.4848
Recall	0.4614	0.1499	0.6326
F1 score	0.4536	0.1971	0.5489
Balanced accuracy	0.5899	0.5047	0.5963
Overall accuracy	0.4434		

Table 8. Random forest model's evaluation metrics on test dataset (without temperature feature) with three classes.

Metrics	Group		
	Adversity	Comfortable	Affluent
Precision	0.4681	0.35610	0.4933
Recall	0.4841	0.19891	0.6307
F1 score	0.4760	0.25524	0.5536
Balanced Accuracy	0.6068	0.53127	0.6034
Overall accuracy	0.4637		

Table 9. Random forest model's evaluation metrics on test dataset (with temperature feature) with three classes.

As with the logistic regression, we examined the performance on the test dataset with two classes. Figure 10 and Table 10 depict our random forest model's performance comparison. It still shows the same outcome that adding temperature can improve accuracy. Compared to the previous logistic model that uses 2 classes, this model's F1 score and accuracy are enhanced.

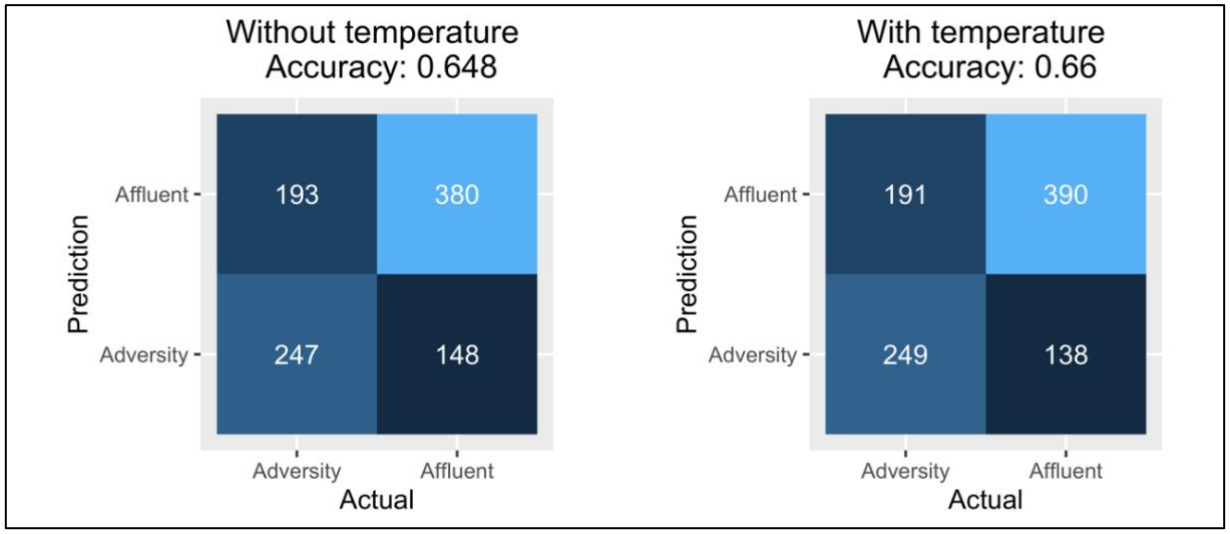

Figure 10. Random forest model's confusion matrices on the test dataset with two classes

Metrics	Without temperature	With temperature
Precision	0.6253	0.6434
Recall	0.5614	0.5659
F1 score	0.5916	0.6022
Accuracy	0.6477	0.6601

Table 10. Random forest model's evaluation metrics on test dataset with two classes.

XGBoost

In addition to random forest, another ensemble approach was utilised. Figure 11 indicates an increase of correct predictions of comfortable and affluent classes on the test dataset that consolidates temperature features while the adversity group's correct prediction slightly decreases. Table 11 and Table 12 explain our XGBoost model's characteristics. All Comfortable group's performance indicators have been improved by using temperature features models. However, the adversity group's recall slightly decreases. As the other indicators generally rise, temperature variables in this model generally contribute to accuracy improvements.

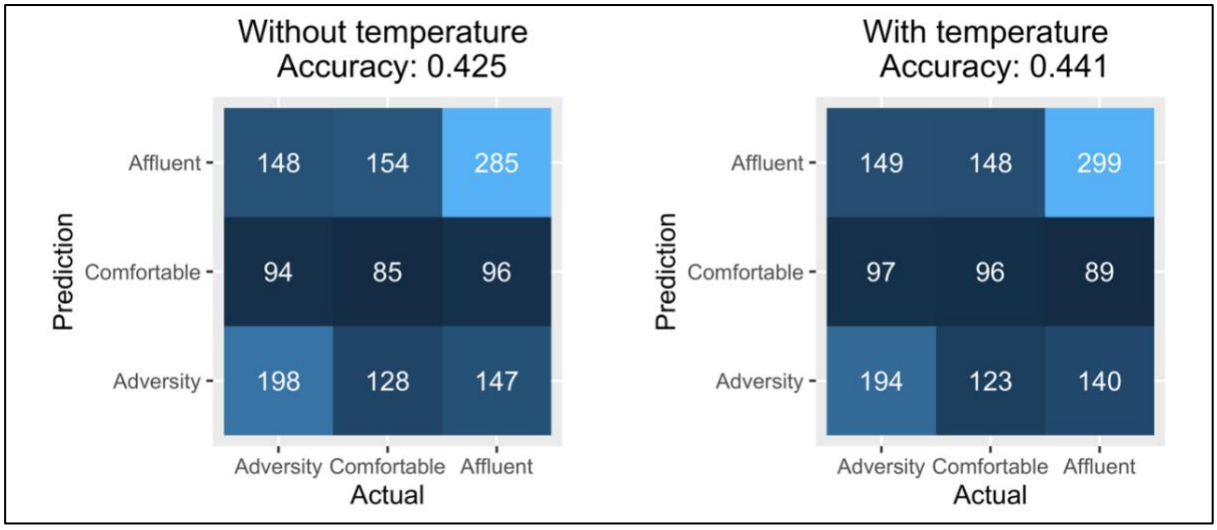

Figure 11. XGBoost model's confusion matrices on test dataset with 3 classes

Metrics	Group		
	Adversity	**Comfortable**	**Affluent**
Precision	0.4186	0.30909	0.4855
Recall	0.45	0.23161	0.5398
F1 score	0.4337	0.26480	0.5112
Balanced accuracy	0.6012	0.51317	0.5818
Overall accuracy	0.4255		

Table 11. XGBoost model's evaluation metrics on test dataset (without temperature feature) with three classes

Metrics	Group		
	Adversity	**Comfortable**	**Affluent**
Precision	0.4245	0.34043	0.5017
Recall	0.4409	0.26158	0.5663
F1 score	0.4326	0.29584	0.5320
Balanced accuracy	0.5899	0.5047	0.5963
Overall accuracy	0.4412		

Table 12. XGBoost model's evaluation metrics on test dataset (with temperature feature) with three classes

When observing the dataset with two classes shows increasing correct predictions, as seen in Figure 12. We can also infer from Table 13 that incorporating temperature information when training the model leads to higher precision, recall, F1 score, and accuracy.

Up until this point, our findings suggest that temperature plays a role to level up the model capabilities.

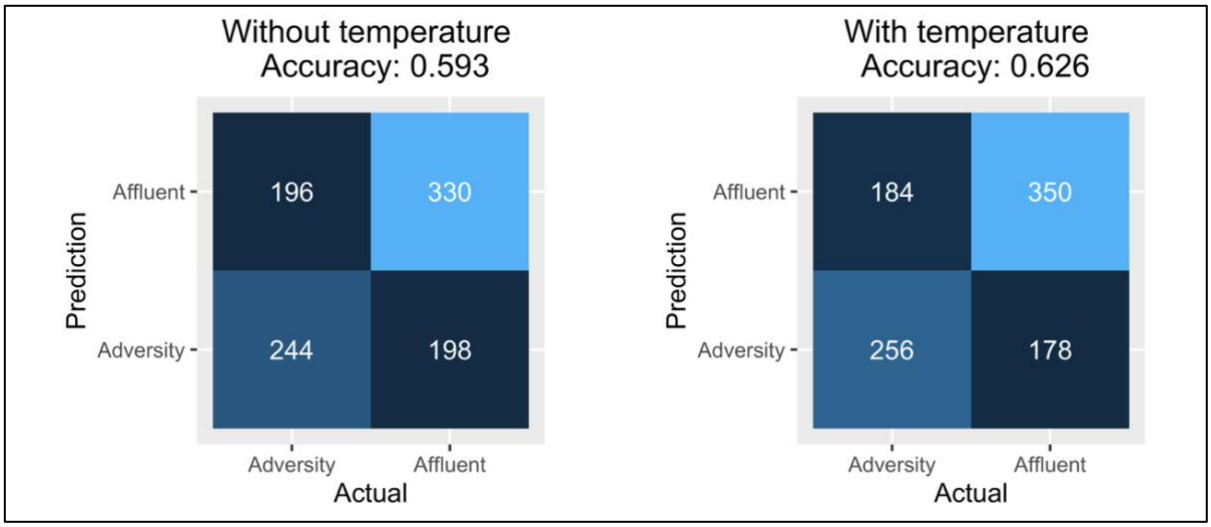

Figure 12. XGBoost model's confusion matrices on test dataset with two classes

Metrics	Without temperature	With temperature
Precision	0.5520	0.5899
Recall	0.5545	0.5818
F1 Score	0.5533	0.5858
Accuracy	0.593	0.626

Table 13. Random Forest model's evaluation metrics on test dataset with two classes

All results

Table 14 highlights surprising facts about the performance comparison of logistic regression and XGBoost. Since these algorithms are distinctive in terms of the way they process, it is interesting to see that these models provide similar accuracy.

Method	Without temperature	With temperature
Logistic Regression	0.43	0.4464
Random Forest	0.4434	0.4637
XGBoost	0.4255	0.4412

Table 14. Accuracy comparison on test dataset with three classes

Table 15 clearly shows the random forest's dominancy when the number of classes is lowered to two. The random forest approach is the only model that can provide accuracy higher than 0.6 on the test dataset without temperature parameters.

Method	Without temperature	With temperature
Logistic Regression	0.595	0.6095
Random Forest	0.6477	0.6601
XGBoost	0.593	0.626

Table 15. Accuracy comparison on test dataset with two classes

5. Conclusions – Summary and Future Work

Following the cyclical, seasonal trend, we have observed a strong reverse correlation between consumption and temperature. The majority of the households were assumed to be without electric heating, based on the mean daily consumption of the dataset. As such, most of the consumption variation is believed to be attributed to the use of electricity for secondary heating and change in lightning behaviour, as the amount of daylight is correlated with temperature. Moreover, we have observed a general increase in consumption from Adversity to Affluent groups, with most of the anomalously high consumption values attributed to the Affluent group. Furthermore, consumption peaks thicken from Adversity to Affluent groups, leading us to conclude that the Affluent group tends to start using excessive lightning and, in some instances, heating earlier in the seasonal cycle than the Comfortable and Comfortable earlier than Adversity group.

Overall, the Affluent group seems the most sensitive to the temperature change and Adversity the least sensitive. Meanwhile, the Comfortable group shows average values, with the group's mean close to the mean of the whole dataset. The decrease of sensitivity toward the Adverse group shows the willingness of the lower-income groups to conserve energy use despite the variation in daylight hours and temperature and further highlights the difficulty of incentivising the more Affluent households to reduce energy consumption.

All models generally perform better when incorporating temperature features from machine learning models' outcomes. Although some models had a slight performance decrease, temperature variables tend to provide higher accuracy. The Comfortable group is relatively hard to be classified as this group's performance metrics are consistently lower than the other two groups' metrics. It became clear that when we dropped the Comfortable class and training the model with only 2 classes – Adversity and Affluent, that all classification metrics were significantly improved. The observed increase of consumption to temperature sensitivity with increased affluence is not translated to the model accuracy. Accuracies of the affluent class are similar or even lower than the accuracies of the adversity class. This stays true for the 2-class case, despite the overall accuracy increase.

Household energy consumption accurately will become increasingly crucial towards a net-zero future. This is evident in the results and method outlined where incorporating temperature improved our model performance. Within the Acorn Categories, there are 17 groups that further break down the demographics of the household. Further studies could be done looking at these groups and how energy is used within them.

Adding data on sunlight and daylight might prove beneficial as the majority of the consumption variation is most likely related to the amount of lightning used in the household. Moreover, filtering the dataset for households with daily consumptions consistently over 10 kWh could improve the model accuracy. The high consumption would be associated with electricity used for primary or secondary heating and, therefore, be more sensitive to temperature changes.

References

Amiruddin, B.P., Kore, E.A., Ulhaq, D.A., Widhatama, A., 2020. Comparison of Classification Algorithms on Household Electricity Consumption Data. International Engineering Students Conference 2020.

Ardakani, F. and Ardehali, M., 2014. Long-term electrical energy consumption forecasting for developing and developed economies based on different optimised models and historical data types. Energy, 65, pp.452-461.

Bonetto, R. and Rossi, M., 2021. Machine Learning Approaches to Energy Consumption Forecasting in Households. [online] arXiv.org. Available at: <https://arxiv.org/abs/1706.09648> [Accessed 12 December 2021].

Brownlee, J., 2021. A Gentle Introduction to XGBoost for Applied Machine Learning. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/> [Accessed 12 December 2021].

CACI. 2020, WHAT IS ACORN?, Available at: https://www.caci.co.uk/sites/default/files/resources/Acorn%20User%20Guide%202020.pdf

Campillo, Javier & Wallin, Fredrik & Torstensson, Daniel & Vassileva, Iana. (2012). Energy demand model design for forecasting electricity consumption and simulating demand response scenarios in Sweden.

Chen, T. and Guestrin, C., 2016. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics, pp. 1189–1232.

Financial Times. 2021. How much will it cost the UK to reach net zero?. [online] Available at: <https://www.ft.com/content/b02b9d51-3e0c-435c-9b53-774ee12ea277> [Accessed 12 December 2021].

Ft.com. 2021. New taxes likely as part of UK's net zero transition by 2050, Treasury warns. [online] Available at: <https://www.ft.com/content/db22c007-6d88-4771-b059-bf0151d3a46a> [Accessed 12 December 2021].

Gumus, M. and Kiran, M., 2017. Crude oil price forecasting using XGBoost. 2017 International Conference on Computer Science and Engineering (UBMK).

Hadri, S., Najib, M., Bakhouya, M., Fakhri, Y. and El Arroussi, M., 2021. Performance Evaluation of Forecasting Strategies for Electricity Consumption in Buildings. Energies, 14(18), p.5831.

HMG, Department for Business, Energy and Industrial Strategy. 2021. [online] Available at: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/960200/CCS207_CCS0221018682-001_CP_391_Sustainable_Warmth_Print.pdf> [Accessed 12 December 2021].

HMG, Department for Business, Energy and Industrial Strategy. 2021. [online] Available at: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/981921/annual-fuel-poverty-projections-2021.pdf> [Accessed 12 December 2021].

Intertek. 2012. Household Electricity Survey. [online] Available at: < https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/208097/10043_R66141HouseholdElectricitySurveyFinalReportissue4.pdf > [Accessed 12 December 2021].

JIANG, Y., CHEN, X., YU, K. and LIAO, Y., 2015. Short-term wind power forecasting using hybrid method based on enhanced boosting algorithm. Journal of Modern Power Systems and Clean Energy, 5(1), pp.126-133.

Jin, N., Yang, F., Mo, Y., Zeng, Y., Zhou, X., Yan, K. and Ma, X., 2021. Highly accurate energy consumption forecasting model based on parallel LSTM neural networks. Advanced Engineering Informatics, 51, p.101442.

Kiprijanovska, I., Stankoski, S., Ilievski, I., Jovanovski, S., Gams, M. and Gjoreski, H., 2020. HousEEC: Day-Ahead Household Electrical Energy Consumption Forecasting Using Deep Learning. Energies, 13(10), p.2672.

Lever, J., Krzywinski, M. and Altman, N. (2016). Classification Evaluation, *Nature Methods*, 13(8), pp. 603-604.

Nielsen, T.R. and Nørgård, J.S., 2009. Household classification according to electricity consumption, *European Council for an Energy Efficient Economy Summer Study Proceedings.*

Pregibon, Daryl. *Logistic Regression Diagnostics*." The Annals of Statistics, vol. 9, no. 4, Institute of Mathematical Statistics, 1981, pp. 705–24, http://www.jstor.org/stable/2240841.

Pudovkin, M.I., (2004). 'Influence of solar activity on the lower atmosphere state'. *International Journal of Geomagnetism and Aeronomy*, 5(2).

Rodrigues, F., Cardeira, C. and Calado, J., 2014. The Daily and Hourly Energy Consumption and Load Forecasting Using Artificial Neural Network Method: A Case Study Using a Set of 93 Households in Portugal. Energy Procedia, 62, pp.220-229.

Research report for Citizens Advice, 2015, Capturing the findings on consumer impacts from Low Carbon Networks Fund projects, pp.13, Available at: https://www.citizensadvice.org.uk/Global/Migrated_Documents/corporate/capturing-the-findings-on-consumer-impacts-from-lcnf-projects.pdf

Shaikh, Faheemullah and Ji, Qiang, Forecasting Natural Gas Demand in China: Logistic Modelling Analysis (2016). International Journal of Electrical Power & Energy Systems, 77: 25-32., Available at SSRN: https://ssrn.com/abstract=2804506

Singh, S. and Yassine, A., 2018. Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting. Energies, 11(2), p.452.

Svetnik, V., Liaw, A., Tong, C., Culberson, J., Sheridan, R. and Feuston, B., 2003. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. Journal of Chemical Information and Computer Sciences, 43(6), pp.1947-1958.

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. and Ahrentzen, S., 2018. Random Forest based hourly building energy prediction. Energy and Buildings, 171, pp.11-25.

Yan, K., Wang, X., Du, Y., Jin, N., Huang, H. and Zhou, H., 2018. Multistep Short-Term Power Consumption Forecasting with a Hybrid Deep Learning Strategy. Energies, 11(11), p.3089.

Zhang, X., Grolinger, K., Capretz, M. and Seewald, L., 2018. Forecasting Residential Energy Consumption: Single Household Perspective. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA).

Zhou, K. and Yang, S., 2016. Understanding household energy consumption behavior: The contribution of energy big data analytics. Renewable and Sustainable Energy Reviews, 56, pp.810-819.

Appendix: Data Source Link

- London smart meter dataset:

https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

-Household information:

https://www.kaggle.com/jeanmidev/smart-meters-in-london?select=informations_households.csv

- Temperature:

https://meteostat.net/en/station/03772?t=2011-01-01/2014-12-31