Enhancing Residential Energy Prediction with Social Variables

Abstract

This paper investigates the utility of social variables to enhance residential energy prediction in light of a strong correlation between the two. I trained Temporal Fusion Transformer (TFT) and Long Short Term Memory (LSTM) models on gas and electricity consumption data from the IDEAL household energy dataset. These models were evaluated using combinations of weather and calendar data, with and without social variables. The results show that social variables have a positive impact on the TFT's prediction of electricity indicated by MAPE and RMSE metrics. However, a similar benefit was not present for the gas dataset due to its larger variability and collection issues. The LSTM did not display an improvement from the addition of social variables, suggesting limitations in its ability to leverage complex relationships within input data. A comparison of the models is also conducted throughout the paper. The findings hope to facilitate a wider integration of social variables to enhance all types of energy prediction.

Author: Sukhjit Singh
Student ID: 2142393
B.Sc. Artificial Intelligence/Computer Science with an Industrial Year
Supervisor: Dr Shuo Wang
40 Credits
Word Count: 4988

1 Introduction

With climate change becoming an ever-increasing issue, the race to reduce carbon emissions is on. Energy use in buildings contributes to 17.5% of all global greenhouse gas emissions and with residential buildings contributing 62% of this (10.9% of all emissions), there is a great incentive to reduce the enormous contributor [40]. The reason for this is that the sources where households get their energy are big polluters like the coal and gas industries, particularly in lower-income countries due to their limited access to modern energy sources [34]. Accurate energy forecasts can help utilities save costs and reduce energy waste in light of intermittent electricity production, a byproduct of the shift to renewable energy. Household electricity load forecasts can be used by regular energy consumers to better understand their habits and Home Energy Management Systems (HEMS) can reduce grid dependence when coupled with rooftop solar panels and other renewable sources [10]. Residential energy forecasts can enable the development of effective demand-response strategies that can lead to efficient energy use, reducing grid operating costs [48]. Forecasting energy load with long-term weather predictions has been shown to optimise energy use, reducing environmental impact [8]. These forecasts are becoming increasingly valuable as climate change can impact energy consumption due to the response to unpredictable weather shocks [5].

With the increasing availability of smart meter data, more suitable data can be applied to forecasting methods to improve prediction. Smart meter data has already been utilised to forecast loads for residential and commercial buildings and has been used to segment customers into different clusters based on energy usage patterns. This can apply to policy-making and help drive targeted energy efficiency programmes [33][26]. The data from these smart meters has enabled the exploration of dwelling and socio-economic data on residential energy use. A thorough literature review of 62 factors that may affect energy use was conducted by [20], including 13 socio-economic and 12 dwelling factors. It finds that the number of occupants, presence of teenagers and household income increase residential electricity consumption. Several factors were inconclusive, such as the homeowner's education level and the presence of elderly people. It also found that bigger and older homes also tend to use more energy. Another study analyses dwelling and occupant socio-economic variables such as number of occupants, home type and build year to observe how they impact household energy use. Among the findings confirmed by the other studies in the previously mentioned literature review, the study finds that apartments use less electricity than houses and again states higher earning professionals consume more than other socio-economic classes [32]. Despite the strong evidence linking socio-economic data to residential energy use, the extent of utilising this data to assist energy predictions is very limited. It is this gap in research that this paper aims to address.

Many machine learning approaches to short-term load forecasting exist with different levels of success [3]. The most effective methods use deep Artificial Neural Networks (ANN) algorithms, particularly Long-Term-Short-Memory (LSTM) models [35][14][24]; however recent studies investigate a transformer architecture [43].

A variant of a transformer, a Temporal Fusion Transformer (TFT), overcomes the traditional transformer architecture and LSTM's limitations by incorporating sequence-to-sequence and attention-based temporal processing units that factor relationships at different timescales. TFTs include a specific network for static variables: static covariate encoders. These encoders learn context vectors from static metadata and deploy them at different stages across the TFT. The self-attention mechanism helps the model to find temporal patterns for specific parts of input data which allows the model to capture long-range dependencies. In relation to energy, this could refer to how energy use in the summer months is lower than winter months. The multi-head attention mechanism takes this further by exploring the patterns for different input variables and identifying more complex patterns in the data [28]. For example, the model could identify energy use is lower in the summer months when there are hotter days and less rainfall. It could learn this relationship and given an unusually hot and dry day in the earlier month of February, more accurately predict the energy use considering the energy use from summer months where conditions are similar. TFTs have seen applications in predicting photovoltaic power and wind speed prediction to help with the renewable transition [29][46].

LSTMs do not offer these features. LSTMs process data sequentially where each step's output depends on the previous steps. This processing does not allow the model to capture long-distance relationships as well as attention mechanisms can. They are less flexible with contextualising different input variables. LSTMs can be adjusted to incorporate many of TFT's features however they require careful engineering to do so.

In this study, I aim to investigate the effectiveness of social variables in improving residential energy forecasts. From the strong evidence and rational intuition of the relationship between socio-economic factors and energy use, I hypothesise that the addition of social variables will improve residential energy forecasts.

In section 2 I will detail my choosing of the TFT and LSTM models to test this hypothesis. Section 3 will feature a breakdown of both models, explaining their components. Section 4 will illustrate the method used to obtain the results in section 5. A discussion of these results will be conducted in section 6 followed by a conclusion in section 7.

2 Related Works

2.1 Energy Forecasting Models

To make short-term energy forecasts, a plethora of statistical and machine-learning models have been used in existing literature with most focusing on the LSTM model due to its observed superior performance [47][31]. Ensemble models have been used to leverage the strengths and weaknesses of multiple models such as ARIMA, neural networks and support vector regression however these require extensive engineering [38][21]. In addition to providing superior performance across many metrics, LSTMs show improved performance when other variables are considered. Study [45] applies an LSTM to predict residential load forecast where they consider weather features as inputs which the study illustrates as a flaw in previous research. The study states that weather features improved their LSTM prediction by 9.87 %.

TFTs have a relatively low application to residential energy use due to their recent creation by Google in 2021 however they have displayed a superiority to LSTMs [28]. Smart homes' energy use was predicted with an LSTM and a TFT across daily, weekly and monthly forecasting horizons [36]. The results conclude that the TFT outperforms an LSTM in MSE, RMSE and MAE metrics across all 3 forecasting horizons as its advanced architecture is better equipped to learn more dynamic relationships.

The literature suggests LSTMs are the established benchmark model in energy forecasting and recent works suggest TFTs' superiority. They both show promise at learning relationships between input features, potentially improving their performance with socio-economic variables, making them ideal models for this experiment.

2.2 TFT

Since TFTs are novel, ordinary transformers have seen a wider application in short-term load forecasting. All of them propose a slightly different version of the transformer, diversifying in input features and prediction hierarchy. A summary of transformers (Trans) and TFTs used for energy forecasting can be found below. From the literature, there is a trend of suggesting an alternate architecture to improve performance rather than conducting thorough tests with the existing models [12]. Identifying other areas to improve such as feature engineering may be a wiser approach to improving overall performance rather than the complex restructuring of model architecture. Smarter feature engineering like selecting weather and calendar data has already been shown to improve energy predictions [9]. Many of these studies do not attempt to illustrate the impact of the features and mostly set out to provide the best possible prediction for their new model. I will use study [12] as a template to follow to show the impact of my feature selection. Of the transformers that have been applied to households, neither apply calendar or weather data to enhance predictions [52][36].

Ref.	Model	Cal	Weather	Social	Level
[28]	TFT				Grid
[30]	Trans		x		Substation
[53]	Trans				Grid
[17]	TFT	x	x		Grid
[16]	Trans				Grid
[51]	Trans	x	x		Grid
[36]	TFT				Household
[6]	Trans	x	x		Grid
[52]	Trans				Household
[27]	TFT	x	x		Grid
[12]	TFT	x	x		Grid
This study	TFT	x	x	x	Household

X represents if the corresponding feature was used in the training of the model
Level refers to the grid level the model is forecasting

2.3 LSTM

Since LSTMs are well established, they have seen wider adoption in the residential energy space. In research, they are typically combined with a variation of neural networks (NN) to leverage the strengths of the NN and the LSTM. An improved performance was seen when using an LSTM with a recurrent NN than without [15]. A convolutional neural network (CNN) LSTM was used to predict the energy use of a single-house which again improved performance compared to an ordinary LSTM[23] and another study uses a particle swarm algorithm to effectively determine the hyperparameters for a CNN-LSTM model, conclusively improving performance versus other forecasting methods [23][22]. A hybrid LSTM was used to predict the energy use of households in London, however, yet again the study lacks utilising the meter data to leverage smarter feature engineering [49].

2.4 Social variables

When predicting energy forecasts, social variables are considered in a very limited capacity. National social factors such as population, gross domestic product (GDP), gross state product and population were used when forecasting the monthly electricity demand for New South Wales Australia [44]. GDP, population and consumer price index were used to help in forecasting Taiwan's electricity consumption using a variety of linear, non-linear and artificial NN models [37]. Similar national social and weather variables were used to predict the total monthly electricity demand of the residential demand of South Korea [41].

Beyond this, there is no practical example of an attempt to use household dwelling and occupancy information to enhance energy forecast models across all statistical and machine learning approaches despite the evidence that there are strong correlative relationships between them.

3 Forecasting Models

3.1 Temporal Fusion Transformer (TFT)

The TFT builds on other transformers by introducing gated residual networks, variable selection networks and static covariate encoders which are explained in the following sections. TFTs provide predictions by looking at past values y of length k, time dependant unknown inputs z, known inputs x and static covariates s.

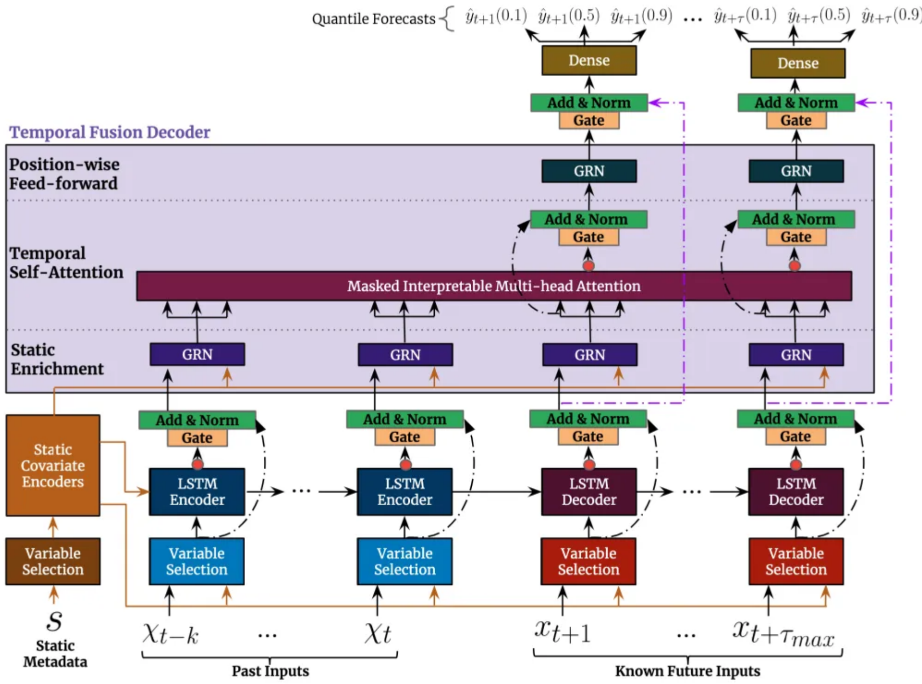

Figure 1: TFT architecture from original paper [28]

3.1.1 Gated Residual Networks (GRN)

GRNs are used to control the flow of information, allowing the model to learn the relevant parts of the inputs and which other parts to discard during the learning process. This improves the generalisation of a TFT, making it flexible for noisy and small datasets. Flexibility is achieved by using several different components. The GRN has two pathways, the residual connection where inputs can bypass any transformations and a transformation pathway where the information is processed and gated. The gating mechanism is a learnable parameter that determines how much of the processed information from the activation function should be passed through the layers. This is important in improving the TFT's ability to capture long-range dependencies and process noise and irrelevant data.

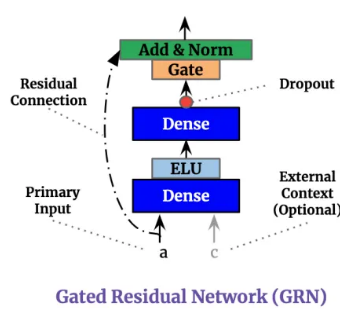

Figure 2: GRN Architecture from original paper [28]

3.1.2 Static Variable Encoders (SVE)

This is a distinguishing feature of the TFT. The SVE helps the model learn context from unchanging, static data. SVEs integrate metadata at multiple layers of the transformer. They influence the initial variable selection and the processing of the temporal information by injecting the static data into their respective layers. SVEs are pivotal in ensuring the predictions from the model do not only reflect the temporal, dynamic input data but static data too.

3.1.3 Variable Selection Networks (VSN)

VSNs are located at the start of the model to help determine which input features are important for the model. This takes place at every time step on a feature level. The TFT learns to weigh the significance of each feature during the training process which the VSN uses to assign a score. The VSN is very dynamic, allowing the model to favour different input features as the time varies.

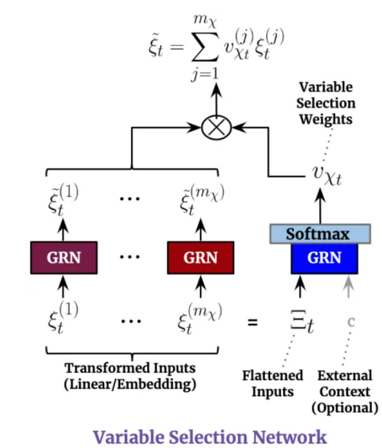

Figure 3: VSN Architecture from original paper [28]

3.1.4 Multi-Head Attention

The attention model is crucial in enabling the model to dynamically focus on different parts of the input sequence, allowing it to learn complex temporal relationships. It does this with three key components:

- Query - The query at the current focus of the attention head. It is the bit of information that the model wants to assess the importance of.
- Keys - Keys act as an index for values. The key is a representation of the value that determines how relevant the value is to the query.
- Value - The value is the actual piece of information the key is referencing. It is the information the model will use in its calculations to learn relationships between the data.

This can be described with the equation

$$\text{Attention}(Q, K, V) = \alpha(Q, K)V \tag{1}$$

Multi-head attention mechanisms take the attention model further by running several attention heads in parallel. It uses multiple sets of queries and keys to capture nuances in the data: W_i^Q, W_i^K, W_i^V where i is the index of the head. First, each attention head undergoes the process described above, measuring the compatibility of the query and keys to compute the attention scores. Once each head has processed the weights, the outputs are concatenated together and passed through a dense layer that is a linear transformation represented by the matrix weight W^O. This whole process can be described with the equation below:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \text{where head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{2}$$

3.2 Long Term Short Memory (LSTM)

The LSTM was designed to address vanishing and exploding gradients that occur during the training of a recurrent neural network. They excel at learning dependencies in a sequence of input data. The input gate, output gate and forget gate work together with the cell state C_t to help learn this information. The cell state acts as memory for the network, carrying information through the sequences to learn information from the data.

3.2.1 Forget State

This gate determines how much of the long-term memory should be forgotten. It determines this by a series of mathematical functions. First, the previous hidden state (short-term memory) h_{t-1} and the input x_t and are multiplied by learned weight matrix W_f. Then a bias b_f is added before feeding the sum into a sigmoid function σ.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

3.2.2 Input Gate

The function of the input gate is to determine what new information will be stored in the current cell state. A candidate layer will create a vector candidate solution $\tilde{C}_t$ that has the potential to be stored in the cell using the following equation, where W_c and b_c are the weights and bias of the candidate layer.

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

The input cell determines how much of this candidate solution should be saved using the following equation where W_i and b_i are the weights and bias for the input layer.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{5}$$

3.2.3 Cell State Update

To update the cell state C_t, first the previous cell state C_{t-1} must forget the amount determined by the output of the forget gate f_t. After this, the new candidate solution is added after scaling by the value from the input gate i_t. This process can be formalised using the following equation.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

3.2.4 Output Gate

After updating the cell state for the next sequence, the output gate determines the hidden state h_t which contains information from the previous inputs. These previous inputs are filtered through the output gate using the equations below where W_o and b_o are the weights and bias for the gate.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t * \tanh(C_t) \tag{8}$$

The entire process can be visualised below.

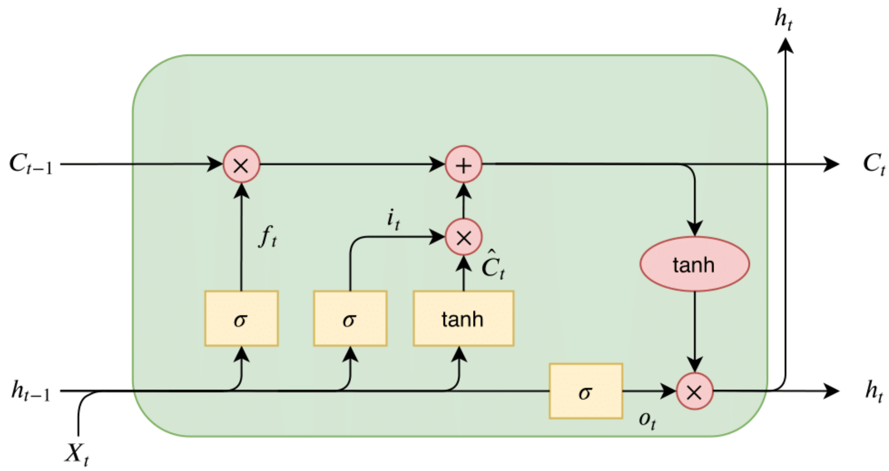

Figure 4: LSTM Architecture

4 Methodology

The IDEAL household energy dataset contains 23 months of detailed gas and electricity data from 255 UK homes recorded for 23 months ending in June 2018. The sensors gathered electricity data by the second and pulse level gas data. Importantly, included in this dataset are the results of a survey that contains property and social information about the household and its occupants respectively. The locations of all the properties are spread across Scotland: Edinburgh, Lothians and South Fife. The included API reduced the difficulty in filtering household data, making it simple to produce the datasets for training [39].

After filtering the household and gas data for all homes, the resulting dataset had to be further reduced and sanitised to be ready for the LSTM and TFT. Due to the practical implications of data collection rollout and connectivity issues, data for all homes is not present for the entire 23 months. The mean number of days of data collected for the homes is 279 days of a potential 654. Because of this, 15 of the homes with the most gas and electricity data available were selected to minimise training time and improve prediction quality. Not all of these 15 homes recorded the same length of data.

With the high resolution of the data recorded from the study, there is freedom over which time step prediction to make. This study will focus on a day-ahead forecast. This type of forecast is long enough to observe patterns in the temporal data, compared to a sporadic hourly forecast [11][4].

When demonstrating if social variables can improve the energy prediction of households, is it important to look at both electricity and gas to get an accurate depiction of the total energy use of homes. In the selected 15 homes, there are 84 gas-powered appliances and 86 electricity-powered appliances so the significance of observing both types becomes apparent due to the similar distributions. Investigating both is essential to determining the impact of social variables.

To evaluate the LSTM and TFT models, I will use mean average percentage error (MAPE) and root mean squared error (RMSE). These metrics are very commonly used across similar studies. MAPE averages the magnitude of errors between predictions and their actual values as a percentage. Its nature as a percentage allows for easy interpretation and its scale independence allows for comparisons across a range of predictions. MAPE is described using the formula:

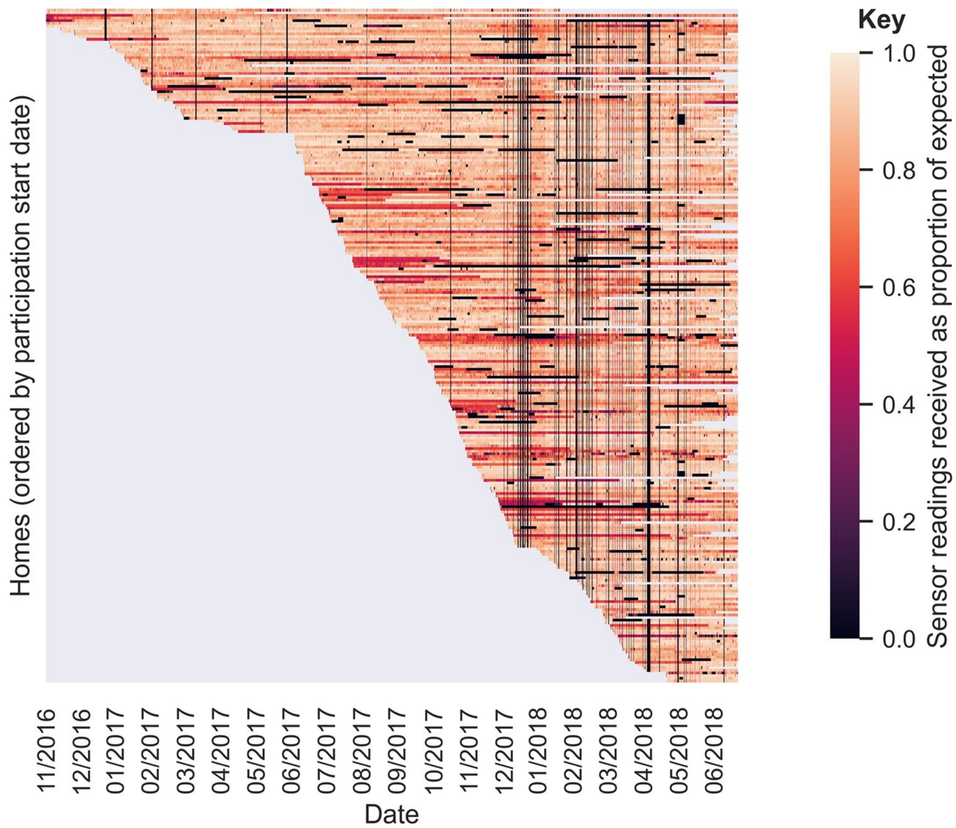

Figure 5: Data received for households in IDEAL Household Energy Dataset

$$MAPE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \times 100\% \tag{9}$$

RMSE is another way of measuring the difference between the predicted and actual values. However, as RMSE averages the squares of the error of the two values, larger errors are emphasised over smaller errors. This offers a counterbalance to MAPE's sensitivity to small values. RMSE can be described with the formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2} \tag{10}$$

In both of these formulas, the inputs can be described by:

- n: the number of observations
- A_i: the actual value for the ith observation
- F_i: the forecasted value for the ith observation

The TFT was built using PyTorch Forecasting's Temporal Fusion Transformer package due to its thorough documentation and flexible implementation. To determine the parameters that will be used by the TFT to obtain a day-ahead energy prediction, I turned to similar studies. Analysing these papers led to the table below where I drew inspiration. The TFT used a learning rate of 0.01, a hidden size of 32, an attention head size of 8, a dropout of 0.1, and a hidden continuous size of 4 with 2 LSTM layers. My input sequences were 100 in length and training used a batch size of 32. A higher learning rate than the reviewed studies was used to reduce training times.

Table 1: TFT parameters from literature - Part 1

Ref	LR	Attention Head	Hidden Size	Dropout	Batch Size
[12]	0.1/0.001	[1-4]	16/32/64	0.1/0.3	32/128/248/496
[17]	0.003		60	0.25	
[29]	[0.001-0.0001]	2/4/6	30/40/60		
[19]	0.001	1	40/60	0.1	72
[36]	0.003	1	16	0.1	1000
This study	0.01	8	32	0.1	32

Table 2: TFT parameters from literature - Part 2

Ref	Hidden Continuous Size	Input Size	LSTM Layers	Optimiser
[12]		[24-672]	1/2/4	
[17]				
[29]	30/40/60		6/4/2	
[19]		72/90	2	Adam
[36]	8	30		MSELoss
This study	4	100	2	Adam

Similarly, I turned to literature for the initial inspiration behind my LSTM hyper-parameters. The hyper-parameters to consider are the layers and neurons in the LSTM, learning rate, loss function, optimiser, dropout and batch size. Most relevant literature that aims to predict household energy using a variant of an LSTM differ in their layer numbers, with 2-3 being a popular choice before a dense layer. The number of nodes in each layer can vary from [2] using 20 but the majority using a value from 100-200. As there is no set rule, I started with a lower value and experimented with higher values. After looking at the data predicted I observed that a value of around 200 with 2 layers captured the trends better than lower values. A higher number of nodes and layers than this did not improve the performance despite requiring significantly longer training times. The learning rate in most literature was the 'adam' optimiser's default parameter setting of 0.001. A variety of loss functions are also used. MSE is used by [13] and [18] with MAE being used by [42]. I will use MSE. Dropout layers are frequently used by the models too, with values ranging from 0.1-0.25. I will use a dropout of 0.1 before the dense layer.

Table 3: LSTM parameters From literature

Ref	Layers & Neurons	Learning Rate	Loss	Type	Optimiser	Dropout	Batch size
[50]	100N x 3L		DWT	SWT-LSTM			56
[2]	20N X 3L	0.001	MAE	CNN-LSTM	Adam		128
[7]	200N X 1L - 100N X 1L		RMSE	Conv-LSTM	Adam		16
[18]	128X3L		MSE	RNN-LSTM	Adam	0.2	32
[42]			MAE	DL-LSTM	Adam	0.25	128
[1]	2L	0.001			Adam	0.1	
[13]			MSE	CNN-LSTM	Adam		64
[25]	20Nx2L	0.001		RNN-LSTM	Adam		
This study	200NX2L	0.001	MSE	LSTM	Adam	0.1	32

An important point to note is that the hyper-parameters will not be tuned beyond the initial values. This is beyond the scope of this paper as the objective is to observe if smarter feature engineering can enhance our prediction, making it essential to control the hyper-parameters.

To optimise the training of the models, reducing my training time was crucial. I used early stopping callbacks provided by the Keras and PyTorch libraries. These allowed me to evaluate the models at every epoch to check if there had been an improvement in performance. If, after a patience of 3, there is no improvement in the evaluation of the model, the callback will be activated and save the best weights of the model. This drastically reduced my training time which is essential given the limited resources of my personal computer.

To investigate the effect of social variables on the performance of the models, I will train them on different combinations of features that belong to 3 categories: calendar, weather and social. Observing the MAPE and RMSE of the models that are trained on different sets of features will help to determine the effect of social variables on the TFT and LSTM models predictions.

Table 4: Summary of Features and Values

Category	**Features**	**Values**
Calendar	Day	[1-31]
	Month	[Jan,...,Dec]
	Year	[2016,2017,2018]
	Day of Week	[Mon,...,Sun]
Weather	Temp	$\mathbb{R}$
	Apparent Temp	$\mathbb{R}$
	Precipitation	$\mathbb{R}$
	Wind Speed	$\mathbb{R}$
Social	Hometype	[Flat,House]
	Residents	$\mathbb{R}$
	Build Era	Before 1850
		1850-1899
		1931-1944
		1945-1964
		1965-1980
		1981-1990
		2002 or later
	Income Band	Less than £10,799
		£16,200 to £19,799
		£23,400 to £26,999
		£27,000 to £32,399
		£32,400 to £37,799
		£37,800 to £43,199
		£48,600 to £53,999
		£54,000 to £65,999
		£66,000 to £77,999

$\mathbb{R}$ denotes the set of real numbers.

Build Era and Income Band have several ranges not fully listed as they were not present in the subset of homes taken from the IDEAL Dataset.

5 Results

This section displays the performance of the TFT and LSTM models using MAPE and RMSE metrics across various feature combinations for electricity and gas data.

Energy type	Features	TFT		LSTM	
		MAPE	RMSE	MAPE	RMSE
Electricity	Power Usage	22.91	2236.61	**26.68**	2157.47
	Social	21.12	1922.26	27.56	2179.62
	Calendar	25.91	2688.94	27.83	**2138.44**
	Calendar+Social	22.38	2057.98	27.97	2141.83
	Weather	18.27	1940.85	27.53	2155.11
	Weather+Social	17.10	1862.09	26.69	2178.64
	Calendar+Weather	16.49	**1662.25**	27.53	2155.11
	Calendar+Social+Weather	**16.17**	1702.85	27.41	2178.64
Gas	Gas Usage	75.51	8151.56	**102.62**	**6832.12**
	Social	78.40	7541.96	129.26	8198.08
	Calendar	78.62	5325.28	111.56	7168.91
	Calendar+Social	85.79	5338.87	115.21	7289.45
	Weather	79.35	7115.53	122.19	7574.36
	Weather+Social	79.93	8779.03	140.95	8043.85
	Calendar+Weather	77.04	6470.66	109.93	7521.08
	Calendar+Social+Weather	**72.13**	**4567.46**	143.20	8136.06

Table 5: Comparison of MAPE and RMSE scores for TFT and LSTM models

These results are averages of 3 runs.
Power and gas usage refer to the historical usage data. The values are included in every set of features.
The best results for each category are in bold.
The models are accessible through a link available in the GitLab repository

MAPE is the average deviation between the prediction and the actual values in percentage terms, so a lower value signifies more accurate predictions by the models. A percentage enables a valid comparison of the performance across datasets. The RMSE is the average magnitude of the prediction errors so a lower value also indicates better performance, however, this metric is dataset-specific. Therefore, comparisons cannot be made between the RMSE for the electricity and gas data but can be made to compare the TFT and LSTM performance within each dataset.

At first glance, the values show that the TFT produces more accurate predictions than the LSTM across all sets of inputs, with or without social features. MAPE scores for the gas dataset are higher than the electricity dataset across both models, suggesting greater challenges with predicting gas usage. The combination of social, weather and calendar features produces the best MAPE score for both electricity and gas datasets. These points will be discussed in the following section. I will compare the models trained on each set of features (historical usage, calendar, weather, calendar+weather) to its counterpart that includes the social features to evaluate the impact of social variables.

6 Discussion

For the electricity prediction using a TFT, there is a notable improvement when social variables are added to the model. From the benchmark model, where the model is trained on only the previous electricity use, adding only the social variables slightly improves the MAPE by 1.79 percentage points (pp) from 22.91% to 21.12% and improves the RMSE from 2236.61 to 1922.26, a 14% decrease. Greater improvements can be seen when social variables are added to the calendar features, the MAPE is improved by 3.53 pp, decreasing from 25.91% to 22.38% and a large RMSE improvement of 23.4%, decreasing from 2688.94 to 2057.98. Social variables also improve the prediction with the weather variables decreasing the MAPE by 1.17 pp and the RMSE by 78.76. Finally, when social variables are added to the model after the weather and calendar variables, the MAPE is the best produced at 16.17% and the second lowest RMSE of 1702.85 of all combinations of input variables.

For the electricity prediction using an LSTM, there is no notable improvement in any metrics from adding the social data. Training the model on the calendar, social and weather data decreases the MAPE by 0.73 pp and increases the RMSE by 21.11 from the benchmark score. Across all types of input features, be it weather, social, calendar or any combination of the 3, the largest improvement in any metric is a 1.28 pp improvement in the MAPE when adding social variables to weather variables. The LSTM's worse metric scores and lack of improvement given input features compared to the TFT is likely due to the TFT's ability to capture relationships between different input data. Exercising the attention mechanisms allows the TFT to better learn relationships between different input variables, improving the overall scores which also aligns with observations found by [17]. The LSTM architecture is good at capturing sequential information yet struggles with finding relationships between the data, therefore limiting the features' contribution to improving MAPE and RMSE scores. The impact of the social variables with the LSTM will naturally be lower than the impact on a TFT. The reason behind this is again in the architecture. TFT's dedicated static-variable encoder enables it to capture and embed static data into the forecasting process. Because of this, the impact of static variables like the socio-economic data will be more pronounced than with an LSTM where, due to the lack of this component, may struggle to leverage static features.

The consequence of social variables is inconsistent for gas predictions using a TFT. Compared to the benchmark, social variables increased the MAPE by 2.9 pp but decreased the RMSE by 609. Adding social variables to the weather increases the MAPE by a marginal 0.58 pp but decreases the RMSE by 12.69. They also worsen predictions for the calendar data, increasing the MAPE by 7.17 pp and RMSE by 13.6. Despite these inconsistencies, social variables improve the MAPE and RMSE when added to both the calendar and weather variables, with the RMSE improvement being vast of 30%, from 6470.66 to 4567.46. This improvement in RMSE could be due to the nature of datasets. The coefficient of variance (CV) is a metric that divides the standard deviation by the mean. This allows for a comparison between the degree of variation of the gas and electricity datasets. The CV for electricity is 0.47 and 1.04 for the gas dataset, meaning the gas data varies more widely around its mean compared to the electricity dataset. Aside from the improvement in the MAPE score, the large improvement in RMSE in this dataset is extremely valuable given the nature of the large swings in gas usage. RMSE is more sensitive to larger errors therefore a reduction implies that social variables are reducing the magnitude of these. The only improvement in both MAPE and RMSE from social data to predict gas comes from adding them to the combination of calendar & weather inputs instead of each variable type separately. This could imply for a dataset with a high CV, more context is required for the TFT to pick on its patterns.

In addition to its high CV, the gas dataset is also of a lower quality which further impacts predictions. Despite best efforts in selecting the cleanest, longest data available, there is still data missing. 20.14% of the data in the selected time was missing for the electricity and 28.86% of the data is missing for the gas dataset. The impact of these gaps on the predictions varies due to their sizes and distributions. Alongside having a higher percentage of data missing, the gas dataset provides further complications due to the distribution of these gaps. The average length of the gaps is 24 consecutive days missing whereas the electricity dataset has an average of 27 days missing per gap in data. The more frequent spread of smaller gaps in the gas dataset makes it difficult for either of these time-series prediction models to produce accurate forecasts. This is why the MAPE scores for the gas dataset are higher compared to the electricity dataset across both models.

The addition of social variables on gas power using an LSTM has an adverse effect. They can be observed to worsen MAPE and RMSE across every combination of input features they are added to. Over the benchmark, social variables increase the MAPE by 26.64 pp and the RMSE by 1365.96. This continues as they worsen MAPE scores with calendar data by 3.65 pp and with weather data by 18.76 pp. The largest adverse effect can be seen when all three variable types are combined, producing a MAPE of 143.20% and RMSE of 8136.06. This outcome is not consistent with the TFT's prediction of the gas dataset, implying that this could be a product of the LSTM architecture's inability to correctly process the static data. Yet, given a marginal difference in the social variables' impact on the LSTM predictions for electricity, it is more likely that the poor quality of the gas data is causing these adverse effects and not the LSTM model itself.

Computational limitations have actualised in only taking 3 runs for each measurement. This has implications as it is not feasible to determine the distribution of the data and therefore it is infeasible to conduct most significance tests. The Wilcoxon signed-rank test does not require a normal distribution however given the few number of samples for each set of input variables, the value of the results would be limited.

7 Conclusion

Socio-economic variables have long been known to influence the energy use of residential households yet there is a clear lack of utilising them for energy forecasts. This paper aimed to explore how valuable this information can be by using smart data from the IDEAL Energy dataset to train a TFT and an LSTM recurrent neural network. The effect of adding social variables is clear for electricity use in the TFT however less clear for the gas data for both TFT and LSTM models due to its low quality and the limitations of an LSTM. The TFT has a dedicated static variable encoder and features attention mechanisms that can capture more complicated relationships than the LSTM which the results support.

A clear and addressed limitation has been the quality of the dataset however there are others to consider. For this paper, a basic LSTM was used whereas in other research, improved LSTM models that have advanced features are typically used. If these superior LSTMs had been implemented, the results could further support that advanced models can enhance their predictions with social variables. Another clear limitation of this paper is the lack of hyper-parameter tuning, as it was to act as a control to observe the impact of the social variables. I intend to tune the parameters for the TFT across different combinations of input features to determine the best MAPE and RMSE scores. In this future study, more measurements will be taken to conduct a valuable statistical test. Once the best values for all sets of features are obtained, I will re-investigate the importance of social variables.

Despite these limitations, this paper has shown that socio-economic variables have the potential to improve household energy forecasts. Improved forecasts can help homeowners optimise their energy efficiency, reducing costs and energy waste. They can also be integrated with the increasing use of renewable energy to reduce the large environmental impact of residential energy consumption on the climate.

References

[1] Derni Ageng, Chin-Ya Huang, and Ray-Guang Cheng. A short-term household load forecasting framework using lstm and data preparation. *IEEE Access*, 9:167911–167919, 2021.

[2] Musaed Alhussein, Khursheed Aurangzeb, and Syed Irtaza Haider. Hybrid cnn-lstm model for short-term individual household load forecasting. *IEEE Access*, 8:180544–180557, 2020.

[3] Ohoud Almughram, Bassam Zafar, and Sami Ben Slama. Home energy management machine learning prediction algorithms: A review. In *Proceedings of the 2nd International Conference on Industry 4.0 and Artificial Intelligence (ICIAI 2021)*, pages 40–47. Atlantis Press, 2022.

[4] Poojitha Amin, Ludmila Cherkasova, Rob Aitken, and Vikas Kache. Analysis and demand forecasting of residential energy consumption at multiple time scales. In *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pages 494–499, 2019.

[5] M. Auffhammer and E. T. Mansur. Measuring climatic impacts on energy consumption: a review of the empirical literature. *Energy Economics*, 46:522–530, 2014.

[6] Yang Cao, Zhenzhen Dang, Feng Wu, Xovee Xu, and Fan Zhou. Probabilistic electricity demand forecasting with transformer-guided state space model. In *2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, pages 964–969, 2022.

[7] Lucia Cascone, Saima Sadiq, Saleem Ullah, Seyedali Mirjalili, Hafeez Ur Rehman Siddiqui, and Muhammad Umer. Predicting household electric power consumption using multi-step time series with convolutional lstm. *Big Data Research*, 31:100360, 2023.

[8] S. Chen, Z. Ren, Z. Tang, and X. Zhuo. Long-term prediction of weather for analysis of residential building energy consumption in australia. *Energies*, 14:4805, 2021.

[9] M. Dahl, A. Brun, O. Kirsebom, and G. Andresen. Improving short-term heat load forecasts with calendar and holiday data. *Energies*, 11:1678, 2018.

[10] Longmao Fan, Jianing Li, and Xiao-Ping Zhang. Load prediction methods using machine learning for home energy management systems based on human behavior patterns recognition. *CSEE Journal of Power and Energy Systems*, 6(3):563–571, 2020.

[11] Nelson Fumo and M.A. Rafe Biswas. Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews*, 47:332–343, 2015.

[12] Elena Giacomazzi, Felix Haag, and Konstantin Hopf. Short-term electricity load forecasting using the temporal fusion transformer: Effect of grid hierarchies and data sources. In *Proceedings of the 14th ACM International Conference on Future Energy Systems*, e-Energy '23, page 353–360, New York, NY, USA, 2023. Association for Computing Machinery.

[13] Xifeng Guo, Ye Gao, Yupeng Li, Di Zheng, and Dan Shan. Short-term household load forecasting based on long- and short-term time-series network. *Energy Reports*, 7:58–64, 2021. ICPE 2020-The International Conference on Power Engineering.

[14] Konstantin Hopf. *Predictive Analytics for Energy Efficiency and Energy Retailing.* PhD thesis, 07 2019.

[15] Tareq Hossen, Arun Sukumaran Nair, Radhakrishnan Angamuthu Chinnathambi, and Prakash Ranganathan. Residential load forecasting using deep neural networks (dnn). In *2018 North American Power Symposium (NAPS)*, pages 1–5, 2018.

[16] Shichao Huang, Jing Zhang, Yu He, Xiaofan Fu, Luqin Fan, Gang Yao, and Yongjun Wen. Short-term load forecasting based on the ceemdan-sample entropy-bpnn-transformer. *Energies*, 15(10), 2022.

[17] Pham Canh Huy, Nguyen Quoc Minh, Nguyen Dang Tien, and Tao Thi Quynh Anh. Short-term electricity load forecasting based on temporal fusion transformer model. *IEEE Access*, 10:106296–106304, 2022.

[18] Nourhan M. Ibrahim, Ashraf I. Megahed, and Nabil H. Abbasy. Short-term individual household load forecasting framework using lstm deep learning approach. In *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pages 257–262, 2021.

[19] He Jiang, Sheng Pan, Yao Dong, and Jianzhou Wang. Probabilistic electricity price forecasting based on penalized temporal fusion transformer. *Journal of Forecasting*, n/a(n/a).

[20] Rory V. Jones, Alba Fuertes, and Kevin J. Lomas. The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renewable and Sustainable Energy Reviews*, 43:901–917, 2015.

[21] Anam-Nawaz Khan, Naeem Iqbal, Atif Rizwan, Rashid Ahmad, and Do-Hyeun Kim. An ensemble energy consumption forecasting model based on spatial-temporal clustering analysis in residential buildings. *Energies*, 14(11), 2021.

[22] Tae-Young Kim and Sung-Bae Cho. Particle swarm optimization-based cnn-lstm networks for forecasting energy consumption. In *2019 IEEE Congress on Evolutionary Computation (CEC)*, pages 1510–1516, 2019.

[23] Tae-Young Kim and Sung-Bae Cho. Predicting residential energy consumption using cnn-lstm neural networks. *Energy*, 182:72–81, 2019.

[24] Weicong Kong, Zhao Yang Dong, Youwei Jia, David J. Hill, Yan Xu, and Yuan Zhang. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1):841–851, 2019.

[25] Weicong Kong, Zhao Yang Dong, Youwei Jia, David J. Hill, Yan Xu, and Yuan Zhang. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1):841–851, 2019.

[26] Jungsuk Kwac, June Flora, and Ram Rajagopal. Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1):420–430, 2014.

[27] Dan Li, Ya Tan, Yuanhang Zhang, Shuwei Miao, and Shuai He. Probabilistic forecasting method for mid-term hourly load time series based on an improved temporal fusion transformer model. *International Journal of Electrical Power Energy Systems*, 146:108743, 2023.

[28] Bryan Lim, Sercan Ö. Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

[29] Miguel López Santos, Xela García-Santiago, Fernando Echevarría Camarero, Gonzalo Blázquez Gil, and Pablo Carrasco Ortega. Application of temporal fusion transformer for day-ahead pv power forecasting. *Energies*, 15(14), 2022.

[30] Alexandra L'Heureux, Katarina Grolinger, and Miriam A. M. Capretz. Transformer-based model for electrical load forecasting. *Energies*, 15(14), 2022.

[31] D. Marino, K. Amarasinghe, and M. Manic. Building energy load forecasting using deep neural networks. 2016.

[32] Fintan McLoughlin, Aidan Duffy, and Michael Conlon. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An irish case study. *Energy and Buildings*, 48:240–248, 2012.

[33] Elena Mocanu, Phuong H. Nguyen, Madeleine Gibescu, and Wil L. Kling. Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks*, 6:91–99, 2016.

[34] M. S. Moeen, A. T. Sheikh, M. Saleem, and S. Rashid. Factors influencing choice of energy sources in rural pakistan. *The Pakistan Development Review*, 55:905–920, 2016.

[35] Sana Mujeeb, Nadeem Javaid, Manzoor Ilahi, Zahid Wadud, Farruh Ishmanov, and Muhammad Khalil Afzal. Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities. *Sustainability*, 11(4), 2019.

[36] Amril Nazir, Abdul Khalique Shaikh, Abdul Salam Shah, and Ashraf Khalil. Forecasting energy consumption demand of customers in smart grid using temporal fusion transformer (tft). *Results in Engineering*, 17:100888, 2023.

[37] Hsiao-Tien Pao. Comparing linear and nonlinear forecasts for taiwan's electricity consumption. *Energy*, 31(12):2129–2141, 2006.

[38] Aida Mehdipour Pirbazari, Ekanki Sharma, Antorweep Chakravorty, Wilfried Elmenreich, and Chunming Rong. An ensemble approach for multi-step ahead energy forecasting of household communities. *IEEE Access*, 9:36218–36240, 2021.

[39] Martin Pullinger, Jonathan Kilgour, Nigel Goddard, Niklas Berliner, Lynda Webb, Myroslava Dzikovska, Heather Lovell, Janek Mann, Charles Sutton, Janette Webb, and Mingjun Zhong. The ideal household energy dataset, electricity, gas, contextual sensor data and survey data for 255 uk homes. *Scientific Data*, 8(1):146, May 2021.

[40] Hannah Ritchie. Sector by sector: where do global greenhouse gas emissions come from? *Our World in Data*, 2020. https://ourworldindata.org/ghg-emissions-by-sector.

[41] Hyojoo Son and Changwan Kim. Short-term forecasting of electricity demand for the residential sector using weather and social variables. *Resources, Conservation and Recycling*, 123:200–207, 2017.

[42] Dabeeruddin Syed, Haitham Abu-Rub, Ali Ghrayeb, and Shady S. Refaat. Household-level energy forecasting in smart buildings using a novel hybrid deep learning model. *IEEE Access*, 9:33498–33511, 2021.

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

[44] D.H. Vu, K.M. Muttaqi, and A.P. Agalgaonkar. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Applied Energy*, 140:385–394, 2015.

[45] Yizhen Wang, Ningqing Zhang, and Xiong Chen. A short-term residential load forecasting model based on lstm recurrent neural network considering weather features. *Energies*, 14(10), 2021.

[46] Binrong Wu, Lin Wang, and Yu-Rong Zeng. Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. *Energy*, 252:123990, 2022.

[47] L. Xiang, T. Xie, and W. Xie. Prediction model of household appliance energy consumption based on machine learning. *Journal of Physics: Conference Series*, 1453:012064, 2020.

[48] C. Xu, C. Li, and X. Zhou. Interpretable lstm based on mixture attention mechanism for multi-step residential load forecasting. *Electronics*, 11:2189, 2022.

[49] K. Yan, W. Li, J. Z, Q. Meng, and Y. Du. A hybrid lstm neural network for energy consumption forecasting of individual households. *Ieee Access*, 7:157633–157642, 2019.

[50] Ke Yan, Wei Li, Zhiwei Ji, Meng Qi, and Yang Du. A hybrid lstm neural network for energy consumption forecasting of individual households. *IEEE Access*, 7:157633–157642, 2019.

[51] Guangqi Zhang, Chuyuan Wei, Changfeng Jing, and Yanxue Wang. Short-term electrical load forecasting based on time augmented transformer. *International Journal of Computational Intelligence Systems*, 15, 08 2022.

[52] Junfeng Zhang, Hui Zhang, Song Ding, and Xiaoxiong Zhang. Power consumption predicting and anomaly detection based on transformer and k-means. *Frontiers in Energy Research*, 9, 2021.

[53] Zezheng Zhao, Chunqiu Xia, Lian Chi, Xiaomin Chang, Wei Li, Ting Yang, and Albert Y. Zomaya. Short-term load forecasting based on the transformer model. *Information*, 12(12), 2021.