# Predicting Household Electricity Consumption based on Weather Conditions and Smart home Appliances

Mansi Saxena; Anushka Kulshrestha; Ravada Naga Venkata Karthik; Shubham

***Abstract*—Smart electricity metering devices advancing by the years allow us to get consumption data on a daily and hourly basis at extraordinary accuracies. With more and more devices making their way into our life our homes have transformed into smart homes. Energy usage and predictability of energy consumption are major parts of sustainable energy usage. Through this project we plan to predict household electricity consumption based on weather conditions and smart home appliances. In this project we are going to implement a predictive analysis of electricity consumption in smart houses based on appliances in the house and the weather conditions of the surroundings using machine learning algorithms such as Linear Regression and neural networks. We will compare the accuracy of the predictions of Linear Regression and neural networks. For our experiment we have taken the data set related to our project from Kaggle. We also identify and predict patterns in the data using data visualisation. The machine learning algorithms have been implemented in python.**

***Index Terms*—Energy Consumed, Energy Generated, Weather Conditions, Smart Home Appliances, Linear Regression, Neural Networks, Recurrent Neural Networks, Long Short Term Memory, Gated Recurrent Units**

## I. Introduction

Throughout the world the interest in smart electricity networks with increased control over electricity supply and consumption has increased considerably. Improvements in smart metering technologies and it is perceived as a necessary step towards achieving sustainable development, to cut greenhouse gas emissions and to improve energy efficiency are the major reason that this was achievable. Electricity usage at the individual household level shows huge fluctuations due to the users' lifestyle, profession, behaviour, building topology and weather conditions. Thus, it is important to analyse and predict the consumption of electricity for analysis of usage and plan for the future. Such usage of machine learning algorithms to analyse the data from smart energy meters to predict the energy consumption of the household is known as sensor-based forecasting approach. The number of studies in energy consumption have seen a drastic increase including annual consumption of various users with the help of advancing and ever-evolving machine learning prediction algorithms and smart metering technologies.

Different studies on this topic have different aims and objectives and come up with a variety of different outcomes. There are also studies that focus on the IT sector. Data centres have become the major need of the time with the rapid evolution of the IT sector. With increasing demand, dependency and usage of cloud-based storage capacity and high data throughput, the most challenging task with such systems is to effectively cool and maintain the systems for better performance and a longer lifetime. There are studies that focus on the relation between weather conditions and energy consumption in Data Centres and show how the weather in the surroundings affect the energy consumption and thus affect the temperature at which the systems operate.

In this paper, we provide a detailed study to answer the following question: how predictable is home energy usage based on the weather of the surroundings? We don't aim to create a new prediction algorithm, but to understand the predictability of home energy consumption with respect to smart appliances in the house and the weather conditions around the house for home energy consumption analysis and management in general. For our analysis, we utilize well-known machine learning algorithms: Linear Regression and Neural Networks. We compare these predictors along with data visualization.

The structure of the paper is organized as follows: a short literature review on similar problems is provided in section 2. In section 3, we provide the preliminaries for the experiment, dataset in section 4, experimental setup in the 5th section and in section 6 will be the analysis. Then, in section 7, there is the proposed architecture.

## II. Literature Survey

[1] In 2016 Lehne introduced the finding of three different scenario of improving access to lighting power and cooking. In order to quantify the patterns of consumer power by deported consumers, a conservation method was used. The required information was obtained from the national statistics on domestic consumers; interviews, archaeological research, research conducted in exile camps. Then, in addition to the results, certain conditions were established. In the first case, migrant households retained their conventional cooking power systems (continued to use the fuel they used previously), but more effectively, while using electricity, homes, previously relied on paraffin and torches, used clear solar lights and diesel generators. In the next case households using solid biomass in cooking have changed about 66 percent of their consumption to biomass briquettes. In the latter case, all immigrant families have used the LPG for cooking and, as in the previous case,

have adopted 40 to 40 subdivisions between solar lights and mini-grid lighting solutions. Comparing the three scenarios, it was said that the first case is easy to see, because the previous costs of these new and powerful technologies are relatively low and the annual fuel savings for foreign buyers could be very significant. The second case will be more costly in terms of fuel costs and capital investments. On the other hand, it has been shown that the latter is more expensive, but will produce more benefits in building a sustainable energy market. As a result, the significant results of the study suggested that for about 7 million migrant buyers in the camps, electricity was provided for less than four hours a day. As a result, the widespread use of state-of-the-art cooking equipment and solar lamps to bring in 303 million dollars in one year requires an investment of 334 million dollars.

[2] In 2015, Muye introduced the conclusion of a study which helps the use of household power by less talented immigrants in rural and urban areas of Bejing. When migrant buyers moved to the host city, they immediately removed biomass from coal, electricity, and LPG. it is fair to say that the energy employed by migrants from rural to urban areas was not exactly the same as domestic consumers, even though all prices were the same. By adjusting from biomass to coal, migrant consumers produced 14 percent more CO2 than urban consumers and a few times .4 of domestic consumers. In the wake of these many rapid changes, migratory domestic energy use patterns show an unprecedented change over the years, as understood in the national home registration system. In terms of electricity consumption, consumer migration patterns are highly dependent on the number of electrical appliances, which are in line with the number of people living during the family period and therefore the time the immigrants lived in Beijing. The amount of electrical appliances and the size of house were not same, indicates that the electrical gadgets were shared by home owners. In contrast, most electrical appliances were purchased by home owners. Therefore, relationships that are directly opposed to the power supply per ca-pita use and house size were noted. in addition, it has been suggested that larger homes should use less energy for lighting, cooking and heating, which means better use, because there are so many different uses. When the shelf life is taken into account, it is found that there is no correlation between the energy consumption and the energy consumption, although the number of appliances was well matched to the residence time.

[3] In a recent study, the primary objective was to explore the potential impacts of migration (urban to urban, rural to urban) on household energy use and carbon dioxide (CO2). The exclusion of a single city in Vietnam, Hanoi (Komatsu et al., 2013). The meaning and significance of Hanoi stems from being a city where migration takes place simultaneously and simultaneously. Using Rosenbaum and Rubin's (1983) scoring system, migrant and domestic consumers are analyzed using economic, human and family-related factors as drivers of power pattern patterns. After all, this has led migrant buyers to consider their connections with various domestic buyers. The general results of the migrant buyer were confirmed in terms of energy consumption and CO2 emissions. The results show that there is no significant statistical impact of urban-to-urban migration in energy consumption and emissions of CO2 per ca-pita, indicating an increase in Hanoi population over urban migration and an increase in the number of people living in rural areas with no significant statistical variance. On the other hand, the migration of consumers from rural to urban areas has also had a positive but negative impact on the overall energy efficiency and CO2 kitchen. In other words, population growth as a result of migration from rural to urban areas leads to lower energy consumption than domestic growth. These findings have critical policy implications for developing urban areas in relation to the relationship between any form of human growth and energy use. For example, policymakers may recognize that the effects of statistical migration are important to patterns of domestic energy consumption of a host city or country. Therefore, it is important for policymakers to consider human capacity while conducting short, medium or long-term energy needs analysis.

[4] Shrestha et al. (2008) reviewed energy efficiency measures involving slum dwellers, Bangkok and Khon Kaen, two Thai cities, depending on energy costs, and important factors affecting access to electricity and other energy resources. In Bangkok, the initial survey was conducted in 2007 in which 100 families were included, this was held in between January and February asked about their social economic status, power generation method, electricity access and other energy source. In addition, these funds test 100 zas made in Khon Kaen houses between September and October 2007. After a study, it was found that almost all households have access to electricity in both cities. About 86 percent of the target group in Bangkok, as well as in Khon Kaen, use oral petroleum gas (LPG) for cooking. They have televisions, refrigerators, rice cookers, washing machines, fans and so on in their homes and in Bangkok, slum dwellers spend about 16 percent of their monthly energy costs and 26 percent in Khon Kaen. For this reason, the study proposes a power plan for these areas, subsidies for slum dwellers and a reduction in service fees in respect of low-income households to increase access to energy services.

## III. Methods and Experimental Setup

As we saw earlier, the main aim of this project is to build a model that predicts the energy consumed by a Smart Home given the weather conditions and the energy usage of various Smart Appliances in the house at one minute intervals. This will help us build a connection between the kind of appliances that we would expect to find in a particular house given the weather conditions there, or the kind of appliances that may be required by a house if we know the other appliances that are being used there. Information like this can be very beneficial to Smart Home Appliance Companies to increase their profits by understanding their general customer's needs better.

TABLE I
TABLE OF LITERATURE SURVEY

0.8— ¿X — ¿X — ¿X— ¿X —

| Study | Objective | Method | Results |
|---|---|---|---|
| Lehne et al. (2016) | Present preliminary estimations for the energy poverty and 3 sophisticated scenarios to improve energy access particularly for cooking and lighting | Present preliminary estimations for the energy poverty and 3 sophisticated scenarios to improve energy access particularly for cooking and lighting | Usage of basic solar lanterns and efficient cooking appliances and will save 303 million dollar/year demanding 334 million dollar investment cost. |
| Muye et al. (2014) | Present survey results about household energy usage of low skilled migrants coming from rural areas in Beijing | Statistical analysis of 1,300 questionnaires | Household energy usage of the migrants doesn't change Too much migrants use different energy mix than residents, but the overall usage amount is similar. |
| Komatsu et al. (2013) | Evaluate impacts of migration on household energy usage and also CO2 emissions in Hanoi | Empirical research, propensity score matching | Policy suggestions concerning energy usage differ urbanization is driven either by migration or by population growth. |
| Shrestha et al. (2008) | Study energy usage habits belonging slum-dwellers in Bangkok and Khon Kaen, as well as associated energy costs, and factors affecting accessing electricity and other energy forms | A survey consisting 100 households between January and February 2007 in Bangkok and the same number of households in Khon Kaen between September and October 2007 | Slum-dwellers spend 16 |
| Fuguitt et al. (1991) | Extend sociological outlook through adding energy consumption behavior analysis | Regression analysis | Metropolitan migrants insert heterogeneity among rural population, due to high energy consumption. |

## A. Methodology

First, we preprocess the data as explained later in part C of this section. Here, the object-type columns of the dataset are converted to float64-type by creating dummy variables or replacing some values. Some rows and columns are dropped to remove all the null or redundant values. The 'time' column is changed to a more suitable format and then made the index. followed by some basic Data Analysis. Then Exploratory Data Analysis is conducted, where various density distribution graphs are made to efficiently compare each column of the dataset to the other. Finally, we split the dataset into the training and testing set in order to prepare it to be applied to various machine learning algorithms.

The first algorithm that we will apply on this dataset is Linear Regression. This is because we are predicting a continuous valued variable ('use [kW]'). We obtain the intercept and coefficient values of the predicted straight line that will pass through our dataset. The Mean Absolute Error, Mean Squared Error, and Root Mean Squared Error along with the score for this model is evaluated.

Linear Regression has some limitations; it requires the data to have some linearity or a linear relationship between the dependent and independent variables. It will not perform well on data that might have other patterns such as cubic or quadruple relationships. However, a Neural Network can gauge these patterns and effectively predict the values of a continuous variable. It has a lot of flexibility in terms of choosing the number of layers and activation units, the regularisation parameter, the optimization algorithm, the optimization algorithm parameters and so on.

Thus, the next algorithm that will be applied on this dataset is a basic Long Short Term Memory Neural Network (or LSTM network). First, we will split the training data into training and validation sets. Then, we build a sequential model of various Dense, fully connected neurons, which are LStm units. This model had 3 layers. The first and second have 32 LSTM units each, while the third layer is a Dense layer with one neuron unit. Two epochs are used, with batch 32 as batch size and validation split as 33 percent. "Mean squared error" is used as the loss function while optimizer is "adam". The accuracy, loss, validation accuracy and validation loss is obtained on each of these sets is plotted to check if the model is overfitting or underfitting the model.

## B. Dataset

The dataset used contains the readings of Energy Consumption from Smart Home Appliances such as the fridge, wine cellar, barn, garage door, microwave, well, furnace. It also contains information about the weather conditions of the area that the house is situated in which are provided with a time span of 1 minute of the Smart Home Appliances in kW from a Smart Meter and weather conditions of that particular region. The dataset contains 503911 records ranging from 0 to 503910 and 32 features. This dataset can be used to understand the relationship that the Energy Consumption of devices and the time period have. By using this dataset, we can understand the relationship between Energy Consumption by Smart Appliances and time period, detect anomalous usage, or clarify the relationship between weather information and energy generated by solar power. We can also check the effect that the kind of weather has on electricity consumption. For this, columns such as temperature, humidity, visibility, apparent temperature, pressure, windspeed, cloud cover, wind bearing, dew point and so on are given.

## C. Experimental SetUp

*Data Fetching and Preprocessing:* The first step in the analysis and finding the relationship between smart home appliances and the energy consumed by them, is to fetch the data of the energy consumed by the Smart home appliances in kW from a Smart meter and the weather conditions of that particular area. The data is stored in a comma separated value (.csv) file. The csv file is read using Pandas library. The analysis done on the Energy and Weather dataset showed that a total of 32 columns were present in the dataset along with 503911 entries. 4 out of 32 columns were of object-data type and the 28 were of float64-data type. The data showed that the last entry, i.e, 503910 had all NaN (Not a number) values and hence is dropped from the dataset. Now, only the non null values remain in the dataset. The time of the dataset is converted from the Unix epoch timestamp to readable DateTime format using Import time and is made the index of our dataset. It can be seen from the above that the time recorded in the dataset is from 2016-01-01 00:00:00 to 2016-12-15 22:29:00.

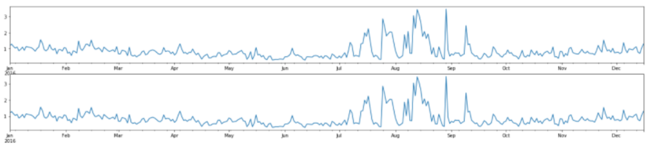

Fig. 1. 'use [kW]' and 'House overall [kW]'

Next, from the figure above, we see that the two columns, namely 'use [kW]' and 'House overall [kW]' are the same. Thus, one of the two columns is dropped to avoid redundancy in our dataset. Then 'icon' column is converted to dummy variables, adding 9 more columns to the current dataset with column names as 'clear-night', 'clear-day', 'rain', 'partly-cloudy-day', 'partly-cloudy-night', 'snow' , 'cloudy', 'wind', 'fog'.Next we replace all the 'cloudCover' values in 'cloudCover' object Datatype with 1 and convert it to a float64 type column. The 'summary' column is also converted to 3 dummy variables 'clear' , 'partly cloudy' and 'cloudy'. At the end of the preprocessing phase we have get a total of 41 columns and 503910 entries with all Non Null values , with 1 column having object dtype, 11 columns having uint8 dtype and the remaining columns having float64 dtype.

### D. Exploratory Data Analysis

In order to visualise and analyse the dataset, we use the python libraries matplotlib, seaborn, holoviews and bokeh. First, the columns 'use [kW]' and 'gen [kW]' are compared. Their density distribution graphs are plotted as "Total Energy Consumption Distribution" and "Total Energy Generated Distribution" vs Density. We see that both are positively skewed. The distribution is very dense around 0.4 kW for the Energy Consumed, and around 0 kW for Energy Generated, ie, it is negligible in most cases.

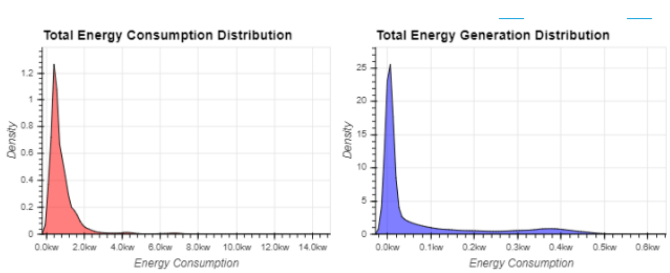


Fig. 2. 'use [kW]' and 'gen [kW]'

Next, the columns that contain information about the energy being consumed by some specific smart appliances in the house are visualised. These columns are "Dishwasher [kW]", "Home office [kW]", "Fridge [kW]", "Wine cellar [kW]", "Garage door [kW]", "Barn [kW]", "Well [kW]", "Microwave [kW]", "Living room [kW]", "Furnace 1 [kW]", "Furnace 2 [kW]", "Kitchen 12 [kW]", "Kitchen 14 [kW]" and "Kitchen 38 [kW]". Only those rows are selected that have the values for these columns greater than 1.5. This is because values greater than 1.5 kW consumption for any appliance in one minute is unlikely, and such data should be treated separately as outliers or exceptions. The density distribution graphs of each column are plotted and compared. We notice that most of these columns have their means very near to 0 kW with Furnace and Living Room as some exceptions. Again, we notice that they are positively skewed.

Lastly, we plot all the columns that contain information about the weather conditions of the area that a particular Smart Home is situated in. These columns include "temperature", "apparentTemperature", "humidity", "visibility", "pressure", "windSpeed", "cloudCover", "precipIntensity", "dewPoint". A density distribution graph of each column is made to provide more clarity. We see that some columns like WindSpeed and Precipitation Intensity are positively skewed, while some like Visibility and Humidity are negatively skewed. Some columns are also similar to standard normal distribution, such as Pressure. We also note that Temperature and Apparent temperature have very similar range of distribution and mean values. Cloud Cover and Dew Point distributions seem to be showing no general pattern as such, and are somewhat erratic. However, we note that the graph for CloudCover suddenly rises at 1. This may be due to the fact that we had replaced the value "cloudCover" with 1 in order to convert this column from object-type to float64-type.

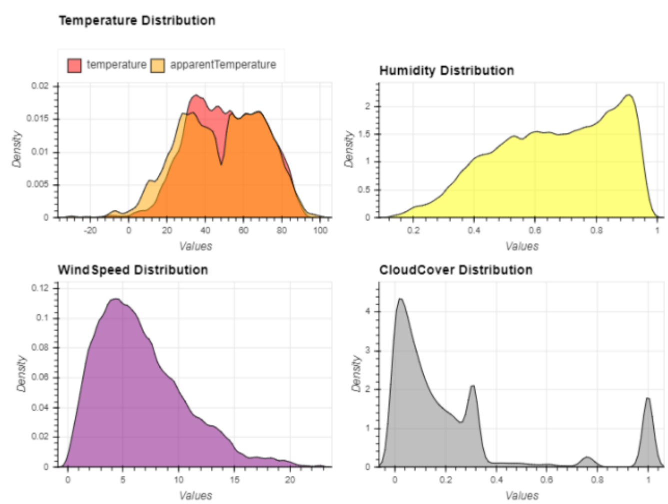


Fig. 3. i) Temperarure; ii) Humidity; iii) WindSpeed; iv) Cloud Cover'

## IV. Preliminaries and Architecture

### A. Preliminaries

Software Requirments:
- Operating System - Windows8 and higher versions
- Language - Python 3.8
- Software - Anaconda, Jupyter Notebooks

Hardware Requirements:
- Ram - 2GB atleast
- Processor - Any Intel processor
- Hard Disk - 8GB and more
- Speed - 2GHz atleast

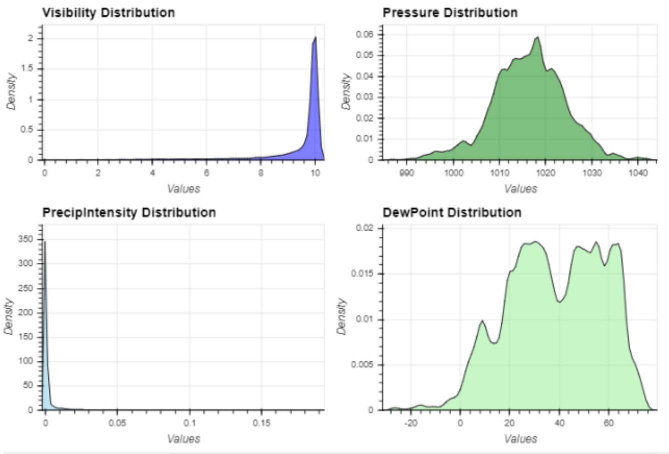


Fig. 4. i) Visibility; ii) Pressure; iii) Precipitation; iv) Dew Point

Python Libraries:

- Pandas - For data manipulation and analysis using dataframes
- Numpy - For high level array and vector mathematics
- Matplotlib - For basic level Data Visualisation
- Seaborn - For advanced Data Visualisation
- Holoviews, Bokeh - For interactive and advanced Data Visualisation
- Sklearn - For various Classification and Clustering algorithms along with performance metrics
- Keras - Framework to implement artificial neural networks

### B. Architecture

*Linear Regression:* The first algorithm that we apply on our processed dataset is Linear Regression. It is a commonly known Supervised Learning - based Machine Learning Algorithm. It predicts or evaluates a dependent variable on the basis of many independent variables or features. In our example here, the dependent variable is the Total Energy Consumed or, 'use [kW]'. The other features, namely the columns that contain information about the weather conditions of the area, the energy consumption by specific appliances and the energy generated, are the independent variables. Thus, the values of these independent variables determine the value of the dependent variable by finding some relationship of pattern between them. These kinds of solutions to supervised learning problems are called Regression models. Various regression models exist based on the number of independent variables under consideration, the relation between these independent variables, the relationship between the dependent and independent variables, and so on.

A linear regression model calculates a regression line that best fits the dataset. This line is a linear function of all the independent variables. All predictions that the model makes are points on this regression line.

Hypothesis function for Linear Regression :

$$Y = \theta_0 + \sum(x_i * \theta_i)$$

Here, $x_i$ are i-number of independent variables, while $\theta_i$ are the i-number of parameters that are trained, that is, the coefficients of the regression line. The most optimized parameter values that are obtained after training the model are the coefficients of the line that best fits the data. $\theta_0$ is the intercept of the line, another parameter that can be trained.

The value that is optimized to make minimum is the root mean square error of the difference between the predicted value and the actual value of the independent variable. This is also called the loss function. The parameters are updated to get the least value of this Loss Function, by an algorithm called Gradient Descent. The values from $\theta_0$ till $\theta_i$ are randomly initialised and Loss function is calculated. Depending on its value, the parameters are updated. This process is called gradient descent and in this way, the regression line is obtained.

*LSTMs and GRUs:* This model is a special type of Recurrent Neural Network to capture patterns in sequential data. Mathematically, a Neural Network is a function that maps one kind of variable to another. Recurrent Neural Networks are specialised Neural Networks to capture text, speech, or sequential data. These RNNs have a major problem of vanishing or exploding values of gradients propagated backward. Gradients are used to update the weights in a neural network. This causes the model to not retain information over long sequences, causing erratic issues to arise when the model trains. This issue can be solved by using an internal mechanism called "gates" that control the flow of information and decide which data in a sequence is to be retained and what can be discarded. It is implemented in the models Gated Recurrent Units (GRU) and Long Short Term Memory Units (LSTM).

In GRUs, a value c known as the memory cell value is used to retain information. This value $c_t$ is updated at each time step t. At every time step, another candidate value, $c'_t$ is calculated along with an update gate value, $u_t$. This update value $u_t$ decides if the value of c must be updated by the candidate value or not. If its value is 1, the value of c is updated, else it is retained. A relevance gate $r_t$ may also be used to determine the importance of the value of c in the previous time step $c_{t-1}$.

Equations of each GRU unit:

$$u_t = \sigma(W_c[c_{t-1}, x_t] + b_u)$$

$$r_t = \sigma(W_r[c_{t-1}, x_t] + b_c)$$

$$c'_t = tanh(W_c[r_t * c_{t-1}, x_t] + b_c)$$

$$c_t = u_t * c'_t + [(1 - u_t) * c_{t-1}]$$

where, $u_t$ is the update gate, $r_t$ is the relevance gate, $hc'_t$ is the candidate for memory cell value and $hc_t$ is the memory cell value at each time step. W and b are the trainable parameters.

LSTMs are more powerful and general than GRUs. There is only one variation in LSTMs from GRUs and that is, $c_{t-1}$ is used in place of $x_t$. It is also known as a peephole connection, and it can be done in any number of gates.

Equations of each LSTM unit:

$$c'_t = tanh(W_c[a_{t-1}, x_t] + b_c)$$

$$u_t = \sigma(W_u[a_{t-1}, x_t] + b_u)$$
$$f_t = \sigma(W_f[a_{t-1}, x_t] + b_f)$$
$$o_t = \sigma(W_o[a_{t-1}, x_t] + b_o)$$
$$c_t = u_t * c'_t + f_t * c_{t-1}]$$
$$a_t = o_t * tanh(h_t)$$

## V. Results and Discussion

By implementing Linear Regression, we obtain Mean Absolute Error as 0.3633188350856295, Mean Squared Error as 0.5186094905867804 and Root Mean Squared Error as 0.7201454648796869. We also obtain the total score of the model as 0.5313002044792383. This shows that Linear Regression does not perform well on our dataset. To understand the reasons behind this, we look at the data points in Figure 5.

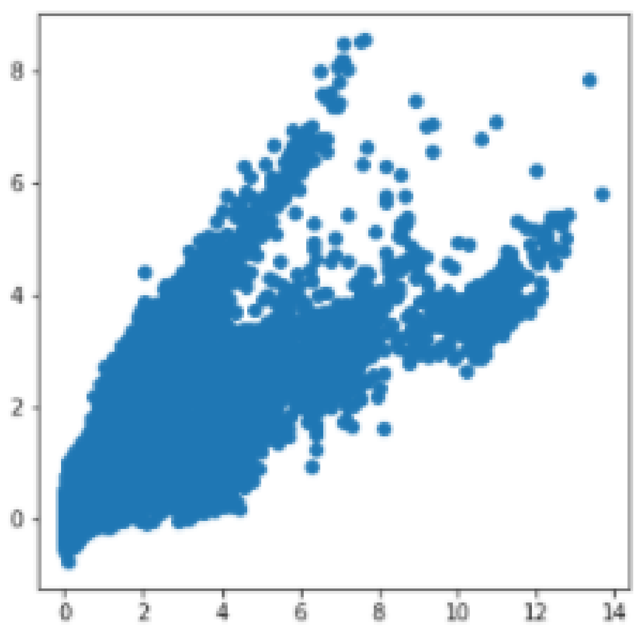


Fig. 5. Plot of the data points

Clearly, the data is not linear. This means that no regression line is capable of giving results with high accuracy on this data. Thus, we use better models, such as GRU and LSTM.

In the implementation of GRU, we use loss as the mean squared error, the adam optimizer, and metrics as accuracy. Training is done for 4 epochs with batch size of 64 and validation split as 33 percent of the training set.

This is the graph that we obtain on implementing our GRU model. Here, blue is the originally used total energy, while red is the predicted usage. The changes in our loss function value, validation loss function value, accuracy and validation accuracy is as follows.

Thus, we see that we have neared a very low value for loss and validation loss, thus showing that GRU performs much better than Linear Regression.

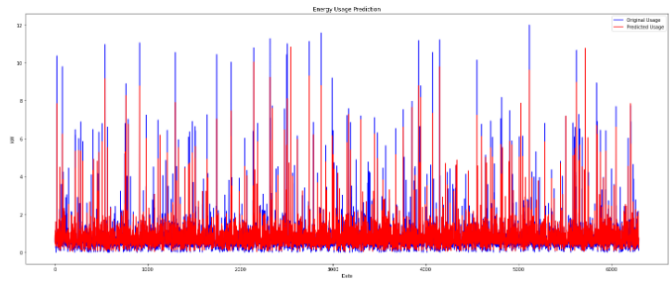

Fig. 6. Plot of actual values compared with predicted values of total energy consumption for GRU

```
LOSS:           [0.9678860902786255, 0.292893566131592, 0.21542569994926453, 0.18596982955932617]
VAL. LOSS:      [0.41466546058654785, 0.22632603347301483, 0.182623952627182, 0.17260104417800903]
ACCURACY:       [8.462635378236882e-06, 1.2693953067355324e-05, 1.2693953067355324e-05, 1.2693953067355324e-05]
VAL. ACCURACY:  [1.7181540897581726e-05, 1.7181540897581726e-05, 1.7181540897581726e-05, 1.7181540897581726e-05]
```

Fig. 7. Results of GRU

In the implementation of LSTM, the same parameters are used as the GRU implementation. This is the graph that we obtain on implementing our LSTM model. Here, blue is the originally used total energy, while red is the predicted usage. The changes in our loss function value, validation loss function value, accuracy and validation accuracy is as follows. These results show that LSTM performs just as well as GRU.

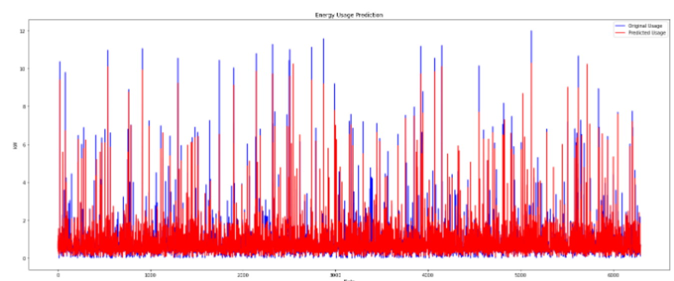

Fig. 8. Plot of actual values compared with predicted values of total energy consumption for LSTM

```
LOSS:           [0.5009031295776367, 0.30108824372291565, 0.24165217578411102, 0.21194612979888916]
VAL. LOSS:      [0.3314262330532074, 0.2528929114341736, 0.19061335921287537, 0.1823696494102478]
ACCURACY:       [8.462635378236882e-06, 8.462635378236882e-06, 1.2693953067355324e-05, 1.2693953067355324e-05]
VAL. ACCURACY:  [1.7181540897581726e-05, 1.7181540897581726e-05, 1.7181540897581726e-05, 1.7181540897581726e-05]
```

Fig. 9. Results of LSTM

However, in practice, LSTMs are said to be more powerful than GRU, especially in the case of extremely large datasets.

Thus, we can conclude that RNN models perform a much better task at predicting the total energy consumed than Linear Regression.

## VI. Conclusion

In this paper we explored the correlation between weather conditions and energy consumption with respect to smart homes and its smart appliances. We obtained real life datasets from Kaggle for this study. The dataset contains the readings of Energy Consumption from Smart Home Appliances such as the fridge, barn, microwave, furnace, etc. It also contains information about the weather conditions of the area that the house is situated in which are provided with a time span of 1 minute of the Smart Home Appliances in kW from a Smart Meter and weather conditions of that particular region the core objective of our study was to make constructive comparison between different machine learning prediction algorithms. We used Linear regression and some Recurrent Neural Network models such as GRUs and LSTMs.

We see it clearly above that the LSTM and GRU models perform much better than the Linear Regression algorithm as they are better suited to produce time series predictions and better estimates of the total active energy consumed , on the basis of the readings from the smart meter. These models are a great tool for anything that has a sequence and a time series. They are also effective at training on the independent features and predicting the values, especially when larger datasets are available. Both these models are able to to extend themselves by creating a short term and long term memory component. This proves very powerful,as they can store the data of previous energy consumed and then learn from it to predict the energy consumed in the next minute or any time asked.

Generally, LSTM models have been noted to be superior to the GRU model when extremely large datasets are under consideration. However, for our implementation here, both work fairly well. This also means that this project can further be enhanced by collecting more information to feed to the neural network and obtain better results.

## References


1 Lehne, J., Blyth, W., Bazilian, M., Grafham, O. (2016).Energy services for refugees and displaced people. Energy Strategy Reviews, 13-14, 134-146.

2 Muye, R., Shu, T., Kirk, S., Guofeng, S., Huizhong S., Ye, H., Han, C., Yilin,C., Xi, C., Junfeng, L., Bengang, L., Xilong, W., Canfei, H. (2015). Direct energy consumption associated emissions by rural-to-urban migrants in Beijing. Environmental Science and Technology, 49(22), 13708-13715.

3 Komatsu, S., Ha, H. D., Kaneko, S. (2013). The effects of internal migration on residential energy consumption and CO2 emissions: A case study inHanoi. Energy for Sustainable Development, 17(6), 572-580.

4 Shrestha, S., Ram, M., Kumar, S. M., Dhakal, A. (2008).Modern energy use by the urban poor in Thailand: a study of slum households in two cities. Energy for Sustainable Development, 12(4), 5-13.S0973-0826(09)60003-X

5 Fuguitt, G., Heberlein, T. A., Rathbun, P. R. (1991). Migration consequences for household energy consumption in a nonmetropolitan recreationretirement area. Rural Sociology, 56(1), 56-69.

6 U. D. of Energy,“Green button,” https://energy.gov/data/green-button.

7 H.-x. Zhao and F. Magoul‘es, “A review on the prediction of building energy consumption,” Renewable and Sustainable Energy Reviews.

8 F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac, “Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines,” International Journal of Electrical Power Energy Systems.

9 Hae Young Noh and Ram Rajagopal. Data-driven forecasting algorithms for building energy consumption, 2013.

10 Richard E. Edwards, Joshua New, and Lynne E. Parker. Predicting future hourly residential electrical consumption: A machine learning case study. Energy and Buildings.

11 Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. Smart*: An open data set and tools for enabling research in sustainable homes. SustKDD, August, 2012.

12 Brad Colman, Kirby Cook, and Bradley J Snyder. Numerical weather prediction and weather forecasting in complex terrain. In Mountain Weather Research and Forecasting.

13 Hargreaves, T., Nye, M., Burgess, J. (2013). Keeping energy visible? exploring how householders interact with feedback from smart energy monitors in the longer term. Energy Policy, 52, 126-134.

14 Nunes, Manuel and Gerding, Enrico and McGroarty, Frank and Niranjan, Mahesan, The Memory Advantage of Long Short-Term Memory Networks for Bond Yield Forecasting (July 6, 2019). Available at SSRN: https://ssrn.com/abstract=3415219 or http://dx.doi.org/10.2139/ssrn.3415219

15 arXiv:1808.03314 [cs.LG]