Household Power Consumption Analysis and Forecasting

Arslan Ahmed
Department of Systems and Computer Engineering, Carleton University
Ottawa, ON
arslan.ahmed3@carleton.ca

Anupam Sehgal
Sprott School of Business, Carleton University
Ottawa, ON
anupam.sehgal@carleton.ca

Olga Baysal
School of Computer Science, Carleton University
Ottawa, ON
olga.baysal@carleton.ca

ABSTRACT

There has been a huge amount of data generated by the power industry in the recent years due to the popularity of smart electrical appliances and home energy management systems . This data can be beneficial for the utility as it provides the behavior patterns of customers, and thus useful decisions can be made to optimize the load on the grid. Our goal for this project is to help the utility companies design better demand side management (DSM) programs to ensure efficient transmission and distribution of energy. This solves the problem of balancing electric demand and supply at the grid and will also reduce the peak demands, which helps lower the electricity bills for the consumers. In this context, we analyze the household electricity consumption data, and determine how it is effected by various factors including consumer's demographics, weather data and dynamic time-of-use pricing. We also develop a forecasting model to predict the energy consumption, and make recommendations to the consumers ahead in time to reduce their load and minimize their monthly bill. To this end, we use Recurrent Neural Networks (RNN) with Long Short Term Memory (LSTM) neurons (also known as LSTM networks) as well as Support Vector Regression with Gaussian kernel. We evaluate their performance using Normalized Root Mean Square Error (NRMSE) metric. We use Low Carbon London (LCL) project dataset available from UK Data Service, London weather dataset from Dark Sky API, and UK bank holidays data from GOV.UK.

KEYWORDS

Smart grid, Energy forecasting, Demand side management, Dynamic Time-of-Use electricity pricing

1 INTRODUCTION

1.1 What is Smart Grids and DS/DR

The idea of a smart grid has been around for a while and recent technological advancement in communications and sensing areas enables the development of smart grid. The traditional power grid landscape consists of centralized generation, where energy is pushed one-way through transmission and distribution networks to the end users. Currently this paradigm evolves by adding bi-directional communication, distributed and utility scale renewable energy generation and energy storage etc. On top of the existing power network layer there is a new communications layer for information exchange and control [1]. Balancing of electric supply and demand at the grid has always been a big challenge for electric utility companies. To counter this, utilities design Demand Response (DR) and Demand Side Management (DSM) programs.

DSM is the planning, implementation and monitoring of utility activities that are designed to influence customer use of electricity. As a result, it changes the time pattern and magnitude of utility's load. Usually, the main objective of demand side management is to encourage users to consume less power during peak times or to shift energy use to off-peak hours to flatten the demand curve. Sometimes instead of flattening the curve it is more desirable to follow the generation pattern. In each case, there is a need of control over customer energy use. [1]
Demand response is a specific tariff or program to motivate end-use customers respond to changes in price or availability of electricity over time by changing their normal patterns of electricity use. It can also be defined as incentive payment program to reduce usage of electricity when grid reliability is jeopardised [2].

1.2 Motivation to apply data science in smart grid

With the increasing penetration of advanced sensor systems in power systems, an influx of extremely large datasets presents a valuable opportunity to gain insight for improving system operation and planning in the context of the electric grids. The 4 Vs of Big Data i.e. volume, velocity, variety, and veracity come with a lot of opportunities as we as challenges [3]. Smart meter data empowers the utilities to make better decisions to optimize the load at the grid. Modern statistical techniques for data exploration enable them to better understand customer's usage pattern, forecast energy consumption, and make recommendations to costumers based on their real-time changing behavior. It also helps them to predict the the downtime and power failures, act proactively, and negotiate about the pricing with the end users. Moreover, it provides the opportunity to the consumers to adjust their loads according to dynamic pricing in order to to minimize their monthly bills [4].

1.3 Challenges

Demand Response and DSM are very promising, however, there are many challenges associated with their implementation.

- Forecasting: This is a challenging problem because of the various uncertainties in electricity peak demand such as population growth, changing technology, economic conditions, prevailing weather conditions. The most challenging part is that we often want to forecast the peak demand rather than the average demand, which varies on various features mentioned above and accurate forecast is a big challenge.

SIGKDD'18, August 2018, London, UK
2018. ACM ISBN 123-4567-24-567/08/06...$15.00
https://doi.org/10.475/123_4

- Customer Profiling: Uncertainties in analyzing and predicting the energy consumption. It is a challenging task to analyze energy consumption of customers with different energy consumption and their will be different peak price recommendation based on different customers.

- Recommendations: Our project seeks to introduce a dynamic pricing system and it is critical to research on how the customers should be recommended so that this new pricing project becomes a success. To obtain a proper recommending medium is a challenging task .

1.4 Our Contributions/Research Questions

In this project, our focus is to optimize the load at the grid, by doing descriptive, predictive and prescriptive data analytics. Our research questions are summarized below. Our objective of this project is to answer these questions.

- How is household electricity consumption affected by the following:
  Consumer's demographics: In our project, due to time constraints, we could research on only one demographic i.e.age category. We analyzed the energy consumption in households with different age categories. This would help in determining the energy consumption and make recommendations to households based on the people living in that house.
  Static characteristics: The energy consumption pattern with respect to different insulation types was analyzed. With the solution to this research question, recommendation could be given to the household regarding which type of insulation type to be used so that they can lower their energy consumption.
- Recommendations to the consumers
  How can customers better respond to changing electricity pricing? : Based on the survey data, asking question to the customers on how they will better respond to the changing electricity rates. This will help in recommending them different aspects with which they will want to take part in the dynamic pricing program, which eventually would help them decrease their electricity usage.
  How can utilities make dToU program more interesting for the residents?: Survey data for the above question would also be effective in knowing about the consumers and recommending on the basis of what they want.
- Forecast short term energy consumption
  What would be the future consumption, given historical consumption and certain features?: This is a critical finding, which would tell how the energy consumption could reduce with respect to different characteristics. There might be certain factors that could change the energy consumption pattern of various households. This will help determine the future trends in energy consumption and maintaining the importance of the project.
  How does the short-term forecasting model perform for long-term forecasting?
  : Using different analytical methods, it is interesting to find out if the short-term forecasting model works well for the long-term also. Does accuracy really improve if we use long period of historical data?: Using long period and short period forecasting data will let us know the different characteristics with which the energy consumption can vary. Using different model it can be also be determined which model is the most appropriate for research in this domain.

1.5 Organization

The rest of the paper is organized as follows. Section 2 provides a review of related work for Data Analytics in Smart Grids. Data sources that are used in this project are explained in Section 3. We describe the methodology for Data Analysis and Forecasting in Section 4. Performance of our forecasting models and other data analysis results are evaluated and presented in Section 5 and 6. Tools used in this project are discussed in Section 7. Finally, the paper is concluded in Section 8 with future work proposed in Section 9.

2 RELATED WORK

In recent years, a plethora of works has appeared in the literature, addressing different aspects of the data analytics in smart grids.In [5], a novel forecasting model for electricity demand time series is proposed. Named the Pattern Forecasting Ensemble Model (PFEM), the new method is based on the pre-existing PSF algorithm, but uses a combination of five separate clustering models: the K-means model (PSF itself), the SOM model, the Hierarchical Clustering based model, the K-means model and the Fuzzy C-means model. Temporal patterns arising in electricity consumption time-series data using a real-world, large-scale dataset have been explored in [6]. A novel algorithm for time series clustering based on Hausdorff distance that efficiently clusters buildings under our distance metric and data stashing technique. The proposed method scales to large data sets, and does not have to be confined to electricity consumption data. Similarly, in [7] proposed method scales to large data sets and is composed of a linear response modeling and a fast heuristic selection algorithm. Another similar and interesting conclusions were found in [8]. This paper provides a simple LP algorithm to be integrated into the EMS of a household or a small business. Via bidirectional communication with the electricity supplier, such algorithm allows maximizing the consumer utility or minimizing its energy cost. The interaction takes place on a hourly basis using a rolling window algorithm to consider the energy consumption throughout the twenty four hours of the day. Bidirectional communication is a key component of a smart grid and as such, is used to design the proposed procedure. A case study demonstrates the usefulness of the proposed algorithm to maximize the utility (or to reduce the electricity bill) of a consumer that integrates the proposed procedure in its EMS. An effective methodology that can be used to drive improvements in peak load forecasting for a power system zone were found in [9]. The proposed approach could inform load forecasting about individual households. Such forecasting is important for design of microgrids and intelligent distribution systems.The methodology suggests that different consumer classes might require different forecasting approaches. In [10], authors proposed a novel framework of collaborative edge-cloud processing for enabling live data analytics in wireless IoT networks. The basic features, key enablers and the challenges of big data analytics in wireless IoT networks have been described and the main distinctions between cloud and edge processing have been presented. Furthermore, potential key enablers for the proposed collaborative edge-cloud computing framework have been identified and the associated key challenges have been presented in order to foster future research activities in this domain

3 DATA SOURCES

3.1 LCL project data

The Low Carbon London (LCL) project was a Â£28m research programme that ran from the beginning of 2011 to the end of 2014 and was funded by energy consumers via Ofgems Low Carbon Network Fund. The programme was designed to investigate the impact of a wide range of low carbon technologies on Londons electricity distribution network. It was in this context that the UKs first residential sector, dynamic electricity-pricing trial took place. This dataset comprises of smart meter readings and consumer survey responses and is available at UK Data Service [11].

Total 5,567 households in the London area participated in the project, of which 1,122 received a dynamic time-of-use (dToU) tariff, and the rest received standard tariff. Data for dToU tariff at different times of the day is also provided. Electricity consumption (kWh) was measured over 30 minute intervals for each household during the trial year of 2013.

Consumer survey responses include data about household characteristics (e.g. type of house, insulation, number of rooms, heating/lightening system, types and number of appliances etc.) and basic details of its occupants (e.g. number of occupants, age categories, gander, etc.). Survey data is available for 1,870 houses with standard tariff and 990 houses with dToU tariff.

Another survey was sent to the houses with dToU tariff to assess their attitudes and behavior change related to dynamic electricity pricing (e.g. change of usage for different types of appliance etc). A total of 714 houses responded to this survey.

It is important to note that all the survey data is categorical i.e. the consumers were given multiple choice questions.

3.2 London weather dataset

It consists of hourly readings of London weather for the year 2013. Various weather parameters are recorded, including but not limited to wind speed, humidity, temperature, pressure, precipitation type (i.e. rain etc), weather icon (i.e. cloudy, sunny etc.) and daily sunrise/sunset timings. This data was taken from Dark Sky API [12]

3.3 UK public holidays data

It consists of public holidays in UK for the year 2013. This data was taken from UK government's official website[13]. Complete featureset dereived from these data sources is described in Table 1.

4 METHODOLOGY

4.1 Data forecasting

4.1.1 Pre-processing. Data pre-processing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects [14]. Data-gathering methods are often loosely controlled,

Table 1: List of features originally available

Half-hourly smart meter readings	Central Heating information	Weather-Humidity
Type of house	Lighting Information	
Insulation type Used	In-Home display	Weather-Apparent temperature
No. of rooms	Survey-Attitudes to low carbon energy	
No. of occupants	Survey-Billing experience	Weather-Wind speed
No. of appliances	Survey-Estimated savings	
Type of appliances	Survey-response to dToU	
Age Category	Survey-improving dToU experience	
Gender	Weather-Temperature	
dToU tariff	Weather-Pressure	

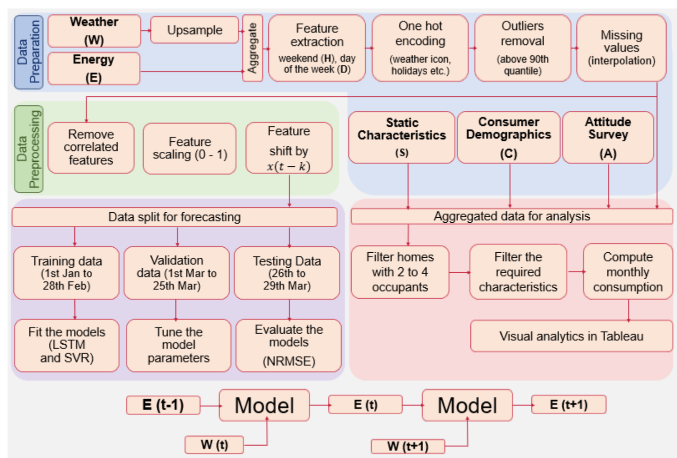

Figure 1: System Model

resulting in out-of-range values (e.g., negative Income), impossible data combinations (e.g., Sex: Male, Pregnant: Yes), missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis [15]. First, we will be preprocessing and cleaning our data, and remove the houses for which we don't have most of the features. Then, we will find correlation between our target variable (energy) and various features, and plot them to get useful insights as outlined in objectives.

- **House Selection:** We filter out the houses based on the following criteria: 1) The number of occupants should be between two and four. This is to maintain consistency in the energy usage patterns because it would be unfair to compare the energy usage of homes with five occupants with the energy consumption of homes with only single occupant. 2)

The homes should have participated in dToU program. This is because one of the objectives of this project was to observe the attitude of the consumers towards similar programs that may be initiated in Canada in near future. After applying these two filters, the total number of houses of interest remain 884 instead of 5567 originally.

- **Re-sampling:** The energy readings are recorded based on half hourly intervals and weather readings are recorded based on hourly intervals, so we first up-sample the weather readings to match its number of rows with energy data. While re sampling, we use the interpolation technique to determine the new value.
- **Combine data tables:** We concatenate two energy and weather tables with Time as an index. This makes it easier to process the data as a whole.
- **Feature Extraction:** We extracted the following three more features out of the existing data: 1) previous i.e. $(t-1)^{th}$ energy consumption value. This is to be stored in LSTM neuron and is controlled by keep gate as explained in section X. 2) Day of the week, starting from Sunday. This is integer value. 3) Boolean variable, Holiday. This uses the holiday dataset and also considers the weekends as holidays.
- **One hot encoding:** Features such as weather icon and day of the week etc. are the categorical features in our table, therefore, we applied one hot encoding to these features. It means that each categorical feature with m possible values will be transformed into m binary features, with only one active. This step is necessary in order to apply our forecasting models as they only expect continuous input and would interpret the categories as being ordered, which is not desired.
- **Outliers removal:** An outlier is an observation that appears to deviate markedly from other observations in the sample. It may indicate bad data, for example, the smart meter may not have been run correctly at that point of time [16]. We observed many outliers in our dataset e.g very high values of energy consumption. These outliers badly effect he scaling of data so we need to remove them beforehand. To remove them, we calculated 90^{th} percentile of the data and deleted all the points above it. Using this method, we made sure that our dataset now represents 90 percent of the original dataset.
- **Handling Missing values:** For various reasons, our dataset had missing values, often encoded as blanks, NaNs or other placeholders. Such datasets are incompatible with scikit-learn estimators which assume that all values in an array are numerical, and that all have and hold meaning. A simple and convenient strategy to use incomplete datasets was to discard entire rows and/or columns containing missing values. However, this comes at the price of losing valuable information. Therefore, we opted for a better strategy imputing the missing values, i.e., to infer them from the known part of the data. For energy consumption values, we took the previous value and used it in place of missing value. For weather data, we used interpolation.
- **Feature selection:** Use extra-tree regressor method to select a subset of features based on contributing weights.
- **Remove correlated features:** We draw a heatmap and check for correlation between features. Correlated features have to be removed otherwise model will overweight them, which is not what we want.
  he final set of eight features that we get are: one lag observation of energy reading, current temperature, Wind Speed, weather-icon, Day-of-the-week, Apparant temperature
- **Feature scaling:** We scale all the features from 1 to 1 using MinMax scaler function of python. This is to make sure that the objective functions assign appropriate weights to all the features.
- **Convert time series into supervised learning problem:** It creates columns of lag observations as well as columns of forecast observations for a time series dataset in a supervised learning format. We use python's shift function.

4.1.2 Applying Models. After pre-processing our data, we apply two different time series regression models and compare their forecasting accuracy.For this, we will split our data into three parts:

- **Training data:** It consists of half-hourly readings between 1st January 2013 to 28th February, 2018. This subset of our data is used to train our models.
- **Validation data:** The validation dataset provides an unbiased evaluation of our model fit on the training dataset while tuning the model's hyperparameters e.g. the number of hidden units in a neural network. It consists of half-hourly readings between 1st March 2013 to 25th March, 2013.
- **Testing data:** It is independent of the training and validation dataset and is seen by the model for the first time. It consists of half-hourly readings between 25th March 2013 to 28th March, 2013. We fit our tuned model over this data to evaluate the results and measure the accuracy.

4.1.3 LSTM networks. Artificial neural networks (ANNs) are based on a collection of connected units or nodes called artificial neurons (a simplified version of biological neurons in an human brain). The signal at a connection between neurons is a real number, and the output of each neuron is calculated by a non-linear function of the sum of its inputs. Weight associated with neurons and connections adjusts as learning proceeds. ANN are organized in layers including input, output and one or mode hidden layers. These layers perform different kinds of transformations on their inputs while the data travels from the first to the last layer [17]. A recurrent neural network (RNN) is a class of artificial neural network where connections between units form a directed graph along a sequence. Unlike feed-forward neural networks, RNNs can use their internal state (memory) to process sequences of inputs [18]. This makes them applicable to tasks that involve sequence such as handwriting recognition and time series predictions. Long short-term memory (LSTM) units (or blocks) are a building unit for layers of an RNN. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell is responsible for remembering values over arbitrary time intervals (which is one in our case); hence the word *memory* in LSTM. Each of the three gates can be thought of as a conventional artificial neuron because they compute an activation of the weighted sum. Building block of LSTM network is shown in Fig. 2 and its internal ANN are described by Eqn. 1 to Eqn. 6.

Algorithm 1 Pseudo code for Data preparation and Data Preprocessing

```
Filter houses which are participating in dToU program using MS Excel
Filter houses which with 2-4 occupants using MS Excel
Read energy data from csv file into Pandas dataframe
Parse the time column as DateTime object and Set the time column as index column
Read weather data from csv file into Pandas dataframe
Parse the time column as DateTime object and Set the time column as index column
Copy data between 1st Jan 2013 to 28th Feb 2014 into two separate dataframes
Resample both dataframes based on half-hourly interval
Concatenate both dataframes based on the index column
Interpolate the dataframe using linear method
Get day of the week, and holiday features and add them as new columns
Encode the new features using One hot encoding
while we reach 10 houses do
    select one house arbitrarily and drop the rest to create a new dataframe
    if the house is not already selected then
        Get an overview of the dataframe using the describe function
        Plot their individual histograms
        Remove the data outside 90^th percentile
        Fill NaN values using pad method
        Sanity check to make sure there is no missing value in the table
        use extra-tree regressor method to select a subset of features that have highest weights
        Plot heatmap of correlation matrix and remove the features that are correlated with each other
        Scale features from 0 to 1 using MinMaxScaler from scikit learn
        convert time series to supervised learning problem and drop the columns that we don't want to predict
    end if
end while
return Pre-processed dataframe for each home, that is fed separately into the model for prediction
```

We use LSTM in our project because they were developed to deal with the exploding and vanishing gradient problem when training traditional RNNs. They are known to be well-suited to classify, process and predict time series given time lags of unknown size and duration between important events. Relative insensitivity to gap length gives an advantage to LSTM over alternative RNNs, hidden Markov models and other sequence learning methods in numerous applications [19].

To implement LSTM, we use Keras on top of TensorFlow. The pseudo code for LSTM networks based forecasting is given in Algorithm 2.

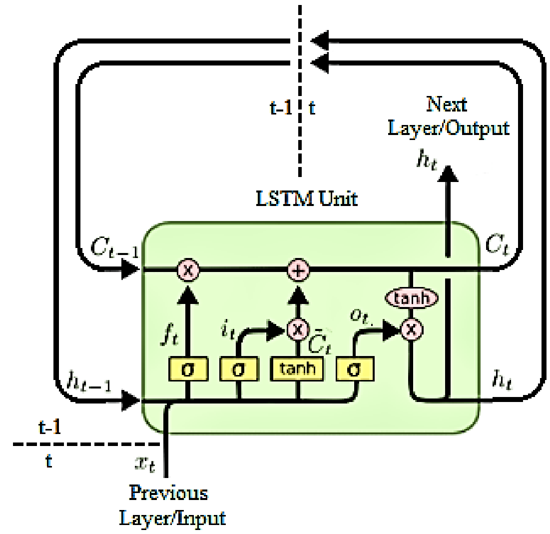

Figure 2: The repeating module in an LSTM Network [20].

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

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2}$$

$$\hat{C}_t = tanh(W_C.[h_{t-1}, x_t] + b_C) \tag{3}$$

$$C_t = f_t \star C_{t-1} + i_t \star \hat{C}_t \tag{4}$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o] \tag{5}$$

$$h_t = o_t \star tanh(C_t) \tag{6}$$

4.1.4 SVR with Gaussian kernel. Support Vector Machine can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences. First of all, because output is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem. But besides this fact, there is also a more complicated reason, the algorithm is more complicated therefore to be taken in consideration. However, the main idea is always the same: to minimize error, individualizing the hyperplane which maximizes the margin, keeping in mind that part of the error is tolerated [21]. Here in this project we use RBF kernel of SVR, which is defined in (7). It takes the form of a radial basis or a Gaussian function.

$$Z(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right). \tag{7}$$

Algorithm 2 Pseudo code for forecasting using LSTM networks

```
for each house's dataframe returned from pre-processed data do
    Set the training data points to 2x24*59 (number of half-hours from 1st Jan till 28th Feb)
    Set the validation data points to 2*24*25 (number of half-hours from 1st March till 25th March)
    Set the testing data points to 2*24*4 (number of half-hours from 26th March till 29th March)
    Split data into three sets i.e. training, validation and testing sets
    Set LSTM parameters according to Table. 2
    reshape the training data array into 3D: [samples, timesteps, features]
    create Keras sequential model to which we will further add core layers
    for 3 iterations do
        Add a layer of LSTM with set parameters
        Add a layer of Dropout to prevent overfitting of NN
        Add a Dense layer which represents the number of output units (1 in our case)
    end for
    Add compile layer to specify the loss function and the optimizer function
    fit the model to the training data
    plot the model loss vs epoch to tune the number of epochs used for training data
    while the loop runs 10 times do
        for each row in validation set do
            reshape the validation row as [samples, timesteps, features]
            predict the next value in the validation set
            add this value to an array of predicted values
            set this value as previous value and increment the validation set row by 1
        end for
        re-scale the predicted array back from 0 to 1 to get original values
        evaluate the error and store it in an error array
        Tune LSTM parameters based on the error of validation set
    end while
    take average of the error array to get NRMSE of one home
    store this error in error_final array
end for
take average of the error_final array array to get NRMSE of 10 homes
return Final NRMSE of our prediction
plot actual vs predicted data
```

$||x-y||^2$ may be recognized as the squared Euclidean distance between the two feature vectors. σ is a free parameter][22]. Pseudo Code for forecasting using SVR is given in Algorithm 3.

4.1.5 Accuracy vs Length of training period. Choosing the length of training period isn't always simple. Its always desirable to get the least error using least amount of training data, however,

Algorithm 3 Pseudo code for forecasting using SVR-RBF kernel

```
for each house's dataframe returned from pre-processed data do
    Set the training data points to 2x24*59 (number of half-hours from 1st Jan till 28th Feb)
    Set the validation data points to 2*24*25 (number of half-hours from 1st March till 25th March)
    Set the testing data points to 2*24*4 (number of half-hours from 26th March till 29th March)
    Split data into three sets i.e. training, validation and testing sets
    Set SVR–RBF kernel parameters according to Table. 2
    while 10 iterations do
        fit the model to the training data
        predict the validation set data
        re-scale the predicted array back from 0 to 1 to get original values
        calculate the NRMSE of one home
        evaluate the error and store it in an error array
        Tune SVR-RBF parameters based on the error of validation set
    end while
    take average of the error array to get NRMSE of one home
    store this error in error_final array
end for
take average of the error_final array array to get NRMSE of 10 homes
return Final NRMSE of our prediction
plot actual vs predicted data
```

there is a trade off between computational expense and the accuracy. Since we are considering short term load forecasting, our goal was to use up to only a few weeks of data for training purpose.
Here in this project we have experimented using 1 to 10 weeks of training data and results are shown in Fig. 4 for LSTM networks as well as SVR using Gaussian kernel.

4.1.6 Accuracy vs Length of forecasting period. We also studied the effect of length of testing data on the forecasting error. We increased the length from one data point to 2*24*4 data points and measured the error at each step, result of which is shown in Fig. 3 for LSTM networks as well as SVR using Gaussian kernel.

4.2 Exploration and Recommendation

4.2.1 Data Transformation and Applying Tableau. Refer to fig.1 for a pictorial view of the exploration. Pre-processing the data has been explained in the previous section. In order to achieve results for the research questions, the data had to be processed in MS-Excel and Tableau in order to obtain adequate visualization for the project. The data had to be processed to an extent that it was easier to apply mathematical formula's and visualization tools. The data had different files for energy consumption, static household characteristics and survey data with common element of household ID. The data had to be transformed so that our research objectives were achieved.The following steps explain our exploration for the

Table 2: Tuning Parameters for forecasting

Parameter	Value	Range checked
batch size	1	1 to 2*24*4
No. of layers on NN	3	1 to 10
No. of neurons in Layer 1	100	1 to 1000
No. of neurons in Layer 2	50	1 to 1000
No. of neurons in Layer 3	20	1 to 1000
No. of timesteps (k)	1	1 to 100
Input shape	1,8	–
return sequences	$true$	–
Dropout	0.2	0.05 to 0.75
Dense	1	–
Loss function	Mean squared error	Mean absolute error and Mean squared error
optimizer	adam	adam, adamax, sgd
epochs	10	5 to 100
shuffle	true	true,false
C	0.1	0.001 to 1
gamma	8	1 to 100

project:

- Data from different files was combined on the basis of the household id.This data was put into Tableau and transformed by applying different operations.
- After applying certain operations, in order to get the average consumption of households based on characteristics such as age and gender. The consumption was realized and visualized using features in Tableau.
- Similarly, data was transformed for survey questions such as alert options that customers found most useful to respond to high/low electricity rate and how to make dToU program more interesting. Aggregation of data and then applied features in Tableau to get the results.

This sums up the exploration part of the dataset.

4.2.2 Recommendation. After exploring the data and finding out the visualizations, it could be easily identified that households with what age or gender consumed the most energy and survey visualizations depicted how can the recommendation be given in future so that the dToU program gains acceptance among the population. The results from exploring the data will be very helpful in implementing the dynamic pricing program and make recommendation to people based on their energy usage and preferences.

5 EVALUATION METRICS FOR FORECASTING

We aim to use Normalized Root Mean Square Error (NRMSE) metric to evaluate our forecasting models. This is a commonly used measure of difference between values predicted by a model and actual observed values.We found the NRMSE for 10 houses with 10 iterations. The following equations were used:

Table 3: Parameters used for calculating NRMSE

Parameter	Description
y_t	Energy at time 't' (actual)
y_t^Λ	Energy at time 't' (predicted)
m	number of observations
y_{max}	maximum energy value
y_{min}	minimum energy value
$NRMSE_i$	error at ithiteration
$\overline{NRMSE_i}$	error of NRMSE of 10 iterations
N	number of houses taken as sample (s)

Table 4: Results

Parameter	Value	Reference
$NRMSE_{LSTM,average}$	0.21	Fig. 5
$NRMSE_{SVR,average}$	0.28	Fig. 5
$NRMSE_{LSTM,1-week-traing}$	0.18	Fig. 4
$NRMSE_{SVR,1-week-traing}$	0.29	Fig. 4
$NRMSE_{LSTM,10-week-traing}$	0.17	Fig. 4
$NRMSE_{SVR,10-week-traing}$	0.23	Fig. 4
$NRMSE_{LSTM,next-point-forecast}$	0.09	Fig. 3
$NRMSE_{SVR,next-point-forecast}$	0.12	Fig. 3
$NRMSE_{LSTM,4-days-forecast}$	0.22	Fig. 3
$NRMSE_{SVR,4-days-forecast}$	0.28	Fig. 3
$AvgConsumption_{all-insulations}$	288.03	Fig. 8
$AvgConsumption_{double-glazing-type}$	154.01	Fig. 8
$AvgConsumption_{adults+kids}$	200.90	Fig. 9
$AvgConsumption_{adults+elders}$	223.45	Fig. 9

$$NRMSE = \frac{1}{y_{max} - y_{min}} \sqrt{\frac{\sum_{t=1}^{m} \left(y_t^\Lambda - y_t\right)^2}{m}} \tag{8}$$

$$FinalNRMSE = \frac{1}{n} \sum_{j=1}^{n} NR\bar{M}SE_i, i = 10 \tag{9}$$

The meaning of the parameters used have been defined in Table 4. Normalizing the RMSD facilitates the comparison between datasets or models with different scales and a low value indicates less variance. We have tested out the value 10 times of 10 different households to get an appropriate result that would determine the accuracy of our forecasting.

6 RESULTS

Following are the results based on our analysis of energy consumption:

- LSTM network perform slightly better than SVR for in forecasting of energy consumption. It can be seen in Figure 5.

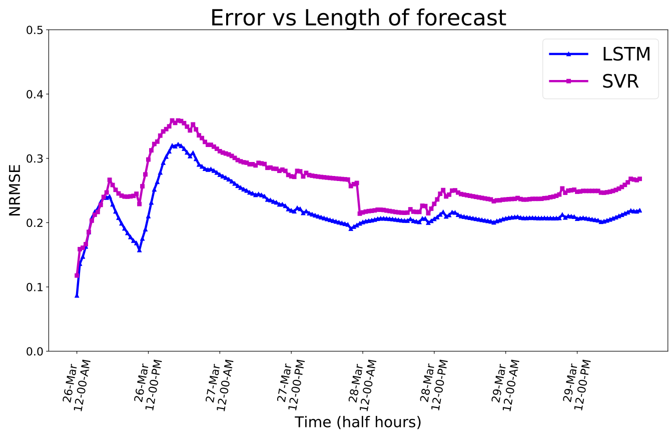

Figure 3: Error vs period of testing data.

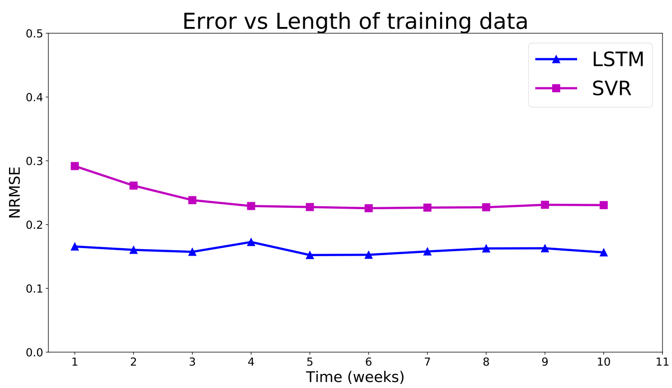

Figure 4: Error vs period of training data.

- In terms of the amount of training data, small amount of data is well sufficient for forecasting which can be understood by figure 4.
- The chances of error increase, when the period of forecasting increases. In figure 3, the NRMSE value is very low and as the length of period increases, the error percentage also increases in both the forecasting models.
- Houses with Wall type insulation consume the highest energy and houses with all types of insulation consume the lowest (see figure 8).
- House with members of age category Adult + Old (18-100 years) consume the most energy in contrast to houses with the age category of Adult + Kids(0 to 64 years). (See figure 9)
- Majority of the customers choose that if the dToU rates are more predictable i.e. they know when the price is going to go up or down, they will be more willing to participate in the dynamic pricing program. (See figure 6)
- Majority of the customers choose that if they are given detailed information on their energy consumption , it will help them respond better to the changing rates and participate in dynamic pricing. (See figure 7)

7 TOOLS

Initially we stored our data locally and used Anaconda and Python Jupyter notebook on Anaconda to process the data. Later as we needed fast processing, we migrated to Google Colab and stored our data on Google drive. Google colab also uses Python Jupyter notebook. Its main advantage is that it allows free access to Google GPUs. LSTM networks were implemented using Keras on top of Tensor flow on Windows 10 machine. Python's scikit-learn, Pandas and numpy libraries were extensively used for data processing.
We used Matplotlib for plotting graphs related to forecasting and Tableau for data exploration and visualization of the static characteristics and attitude survey. MS Excel was also used initially for filtering the houses of interest.
All the graphs presented in this report are from Matplotlib because it was easily customizable. The original graphs were made in Tableau and data was extracted to be used in Matplotlib.

8 CONCLUSIONS AND IMPLICATIONS

We have concluded the detailed analysis and forecast of the household energy consumption. Based on the results, it can be concluded that LSTM neural networks perform better than SVR.Hence, LSTM network is recommended to be used in similar kinds of projects for forecasting. Small amount of training data is ample for accurate forecasting in household energy consumption and it can provide results with a low Root mean square error. With increase in the time period of testing data, the possibility of error will also increase when predicting the future energy consumption. Hence, the forecasting must be done on a short time period to get accurate results. The results on comparison of insulation material used in households conclude that consumers can save up to 30.5 percent energy by using full insulations instead of the commonly used double glazing insulation. Based on the results from analysis on the answers given in surveys, it can be concluded that Utilities can get better result out of dToU program by keeping the consumers well-informed about their consumption, and making the dToU more predictable e.g. every Sunday between a certain time duration, the price per unit will be more than usual. Also, the consumers will better respond to the changing rates when they are given better information on their consumption.

9 FUTURE WORK

Below are some of the tasks that we were not included in the scope of this project due to time constraints. We list them as future work.

- **Remove seasonality of time series:** Very often in the literature we have seen the researchers detrend the time series data before applying any forecasting model.
- **Perform multi-dimensional processing** So far we have been dealing with only two-dimensional data after selecting a subset of original dataset available. We can add various static features as features and perform higher-dimensionality processing which may give us better results.
- **Apply k-means clustering** This will give more effective clusters for dynamic pricing than filtration techniques.

REFERENCES

[1] L. Gelazanskas and K. A. Gamage, "Demand side management in smart grid: A review and proposals for future direction," *Sustainable Cities and Society*, vol. 11,

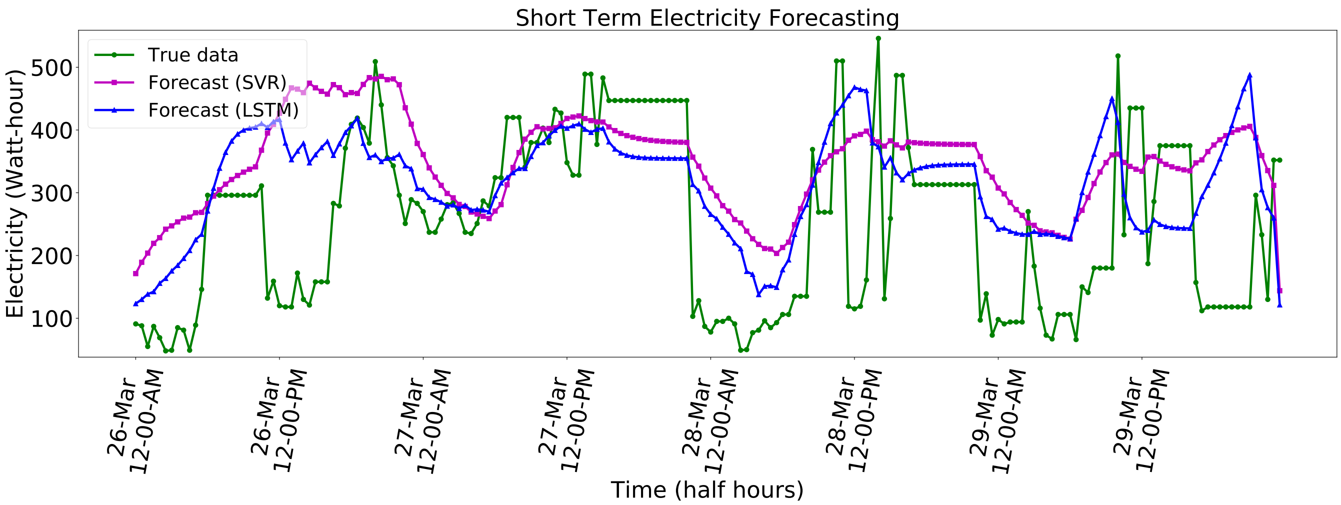

Figure 5: Short Term Electricity Forecasting (LSTM vs SVR).

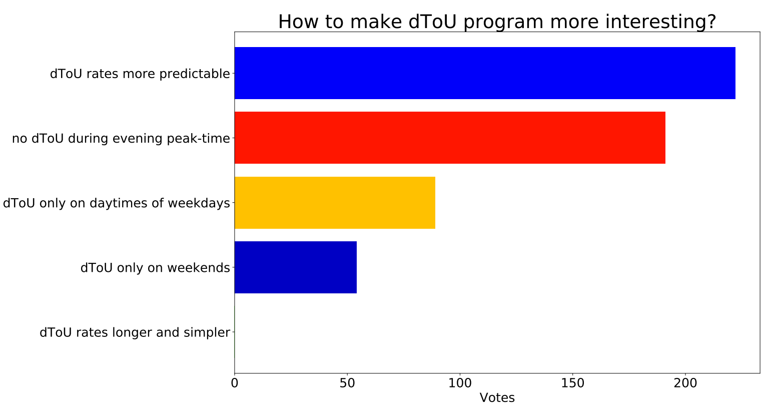

Figure 6: Recommendation to utilities.

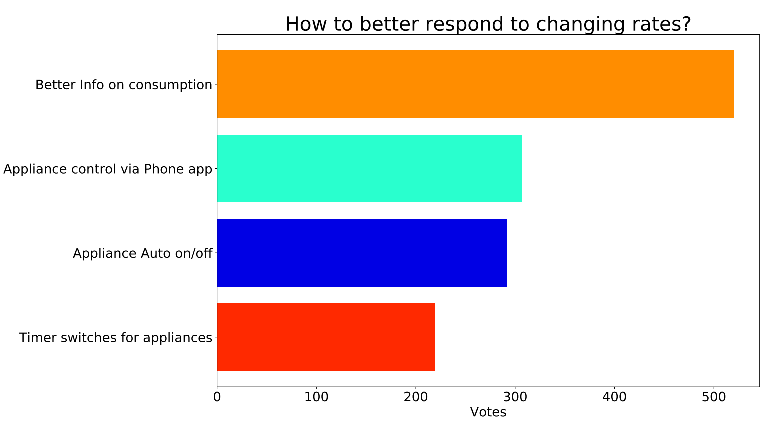

Figure 7: Recommendation to consumers

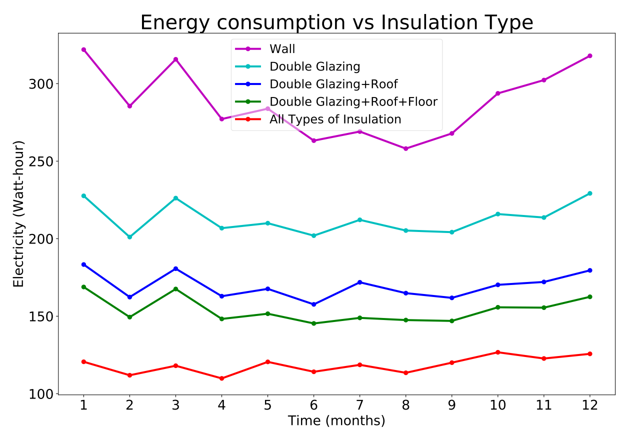

Figure 8: Energy consumption vs Insulation type.

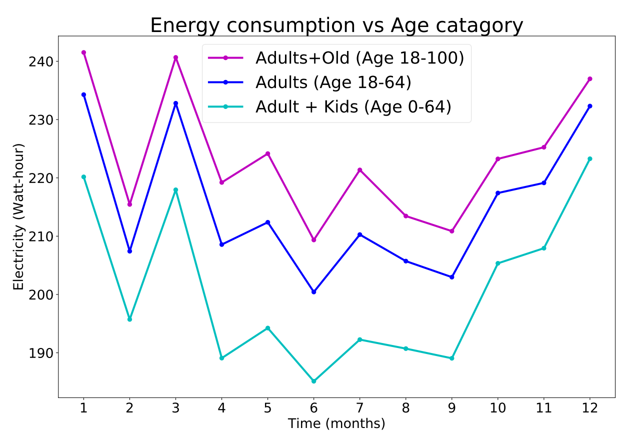

Figure 9: Energy consumption vs age catagory.

pp. 22–30, 2014.
[2] Q. Qdr, "Benefits of demand response in electricity markets and recommendations for achieving them," *US Dept. Energy, Washington, DC, USA, Tech. Rep*, 2006.
[3] "Infographic: The four v's of big data | ibm big data & analytics hub." http://www.ibmbigdatahub.com/infographic/four-vs-big-data. (Accessed on 04/21/2018).
[4] Z. B. A. Ahmed, K. Arab and M. Ibnkahla, "Data communication and analytics for smart grid systems," *Proc 2018 IEEE International Conference on Communications*, pp. 1–6, 2018.
[5] W. Shen, V. Babushkin, Z. Aung, and W. L. Woon, "An ensemble model for day-ahead electricity demand time series forecasting," in *Proceedings of the fourth international conference on Future energy systems*, pp. 51–62, ACM, 2013.
[6] C. Chelmis, J. Kolte, and V. K. Prasanna, "Big data analytics for demand response: Clustering over space and time," in *Big Data (Big Data), 2015 IEEE International Conference on*, pp. 2223–2232, IEEE, 2015.
[7] J. Kwac and R. Rajagopal, "Demand response targeting using big data analytics," in *Big Data, 2013 IEEE International Conference on*, pp. 683–690, IEEE, 2013.
[8] A. J. Conejo, J. M. Morales, and L. Baringo, "Real-time demand response model," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 236–242, 2010.
[9] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420–430, 2014.
[10] S. K. Sharma and X. Wang, "Live data analytics with collaborative edge and cloud processing in wireless iot networks," *IEEE Access*, vol. 5, pp. 4621–4635, 2017.
[11] J. R. Schofield, R. Carmichael, S. Tindemans, M. Bilton, M. Woolf, G. Strbac, *et al.*, "Low carbon london project: Data from the dynamic time-of-use electricity pricing trial, 2013," 2015.
[12] "Dark sky." https://darksky.net. (Accessed on 01/30/2018).
[13] "Uk bank holidays." https://www.gov.uk/bank-holidays. (Accessed on 01/30/2018).
[14] "Data pre-processing - wikipedia." https://en.wikipedia.org/wiki/Data_pre-processing. (Accessed on 04/21/2018).
[15] D. Pyle, *Data preparation for data mining*, vol. 1. morgan kaufmann, 1999.
[16] "Outlier - wikipedia." https://en.wikipedia.org/wiki/Outlier. (Accessed on 04/21/2018).
[17] "Artificial neural network - wikipedia." https://en.wikipedia.org/wiki/Artificial_neural_network. (Accessed on 04/21/2018).
[18] "Recurrent neural network - wikipedia." https://en.wikipedia.org/wiki/Recurrent_neural_network. (Accessed on 04/21/2018).
[19] "Long short-term memory - wikipedia." https://en.wikipedia.org/wiki/Long_short-term_memory. (Accessed on 04/21/2018).
[20] "Understanding lstm networks – colah's blog." http://colah.github.io/posts/2015-08-Understanding-LSTMs/. (Accessed on 04/21/2018).
[21] "Support vector regression." http://www.saedsayad.com/support_vector_machine_reg.htm. (Accessed on 04/21/2018).
[22] "Radial basis function kernel - wikipedia." https://en.wikipedia.org/wiki/Radial_basis_function_kernel. (Accessed on 04/21/2018).