8

Household Power Consumption Forecasting using IoT Smart Home Data

Fitri Indra Indikawati[1], Guntur Maulana Zamroni[2]

[1,2] Department of Informatics, Universitas Ahmad Dahlan, Indonesia

ARTICLE INFO

ABSTRACT

The use of the forecasting system is becoming more prominent in recent years. One of the implementations of the forecasting system is to predict electricity consumption demand. In this paper, we have developed a forecasting system for household electricity consumption using a well-known Extreme Gradient Boosting algorithm. We utilized time-series data from a smart meter dataset to make a predictive model. First, we evaluated the importance of time-series feature from the dataset and resampled the original dataset. Then, we used the resampled data to train the model and calculated training loss function. Our experimental studies with real IoT Smart Home data demonstrate that our forecasting system works well with small dataset using one-hour downsampling on the dataset.

Corresponding author:

Fitri Indra Indikawati,
Department of Informatics,
Universitas Ahmad Dahlan,
Ring Road Selatan, Special Region of Yogyakarta 55191, Indonesia
Email: fitri.indikawati@tif.uad.ac.id

1. INTRODUCTION

Due to the rise of economic growth, industrial development and household living standard are increasing. One of the consequences of this phenomenon is higher electricity consumption [1],[2]. Based on the study of 210 countries over 54 years [3] and 160 countries over 30 years [2], there is a bidirectional correlation between economic growth and the increase of electricity consumption, especially in developing countries.

The economic growth of developing countries usually relies on manufacturing industries. Industries consume an abundant amount of electricity. Slight shortage of electricity can cause a massive decline in production. A decline in production means a financial loss for the industries and the country, hence hindering the economic growth of the country.

However, a surplus of electricity supply also causes another problem. Even though renewable energy resources are becoming more prevalent, most of our energy resource for generating electricity came from non-renewable resources, such as coal or oil. An overestimation in electricity consumption means wasted energy resources. Therefore, planning policy for electricity supply and demand should be implemented to maintain the economic stability of a country.

Electricity consumption modelling plays a vital role in electricity planning policy. Analyzing electricity consumption by using historical consumption data can be done to ensure the effectiveness of energy planning. Electricity forecasting model cannot be generalized because each country has different electricity consumption trend [4]–[8]. By using IoT smart meter, electricity consumption trend can be seen not only in the country level but also in lower level such as city, residential area, until household level [9]–[11].

Due to the rise of IoT device in a smart home, nowadays each household can track electricity consumption by using a smart meter. Using a smart meter, the house owner can see not only overall electricity consumption but also electricity consumption in each room or specific appliances. This data can be used to track electricity usage patterns for specific appliances or room, get a usage estimation based on time range, and to manage and plan to spend on electricity bills.

Data from the smart meter can also be used by the energy companies in each residential area or city to get a better understanding of the electricity demand in that area. Forecasting electricity consumption can help power companies to create more suitable electricity tariffs [11], to understand electricity consumption trends at certain times, or to create efficient energy planning based on supply and demand.

The growing necessity of creating a forecasting model of electricity consumption has led to extensive research on this topic. However, due to the nature of time-series data, creating a forecasting model is not an easy challenge. To build an electricity consumption forecasting system, we investigate a few challenges: 1). To select suitable time-series feature, (2) implement model and algorithm to predict future consumption.

The first challenge is to select a suitable time-series feature to make a forecast. Different time series features are used for different reasons. For annual energy planning, forecasting electricity consumption can be done using yearly time-series [4]. It can also be done using monthly time series [7], daily time series [10] or hourly time series [6]-[8] for a more detailed forecast.

There are several methods to make a predictive model, for example, ARIMA [4]-[6], regression analysis [5], neural networks [5]-[7], and least squares support vector machines [5]. Most of the forecasting methods usually need quite a massive amount of data to train the forecasting model.

To address these challenges, we propose to build an electricity consumption forecasting model by identifying suitable time series feature before making a prediction and using extreme gradient boosting [12] to train our model due to the small dataset size.

The remainder of this paper is organized as follows. In Section 2, we discuss the overall architecture of the proposed system, forecasting model and algorithm. Then, we discuss the experiment results of the system in Section 3. Finally, Section 4 concludes the paper.

2. RESEARCH METHOD

In this section, we present the design of our electricity consumption forecasting system.

2.1. System design

The overall process of our system is depicted in Fig. 1. The process consists of several steps: select time-series feature from the dataset, dataset resampling, and training model. The explanation of Smart Home Dataset can be found in subsection 2.2.1.
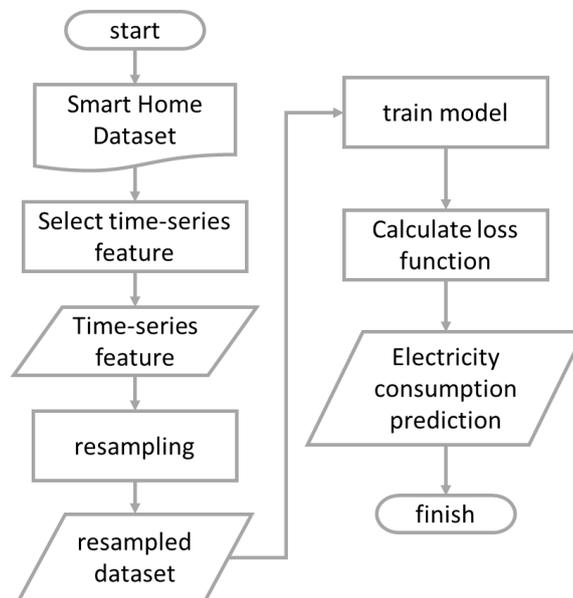

Fig 1. Forecasting system flowchart

First, we create a time-series feature from original dataset and evaluate feature importance for each time-series feature. Then, we use time-series feature with high importance to resample the dataset. The next process is to use the resampled dataset to train the data model using gradient tree boosting. We utilize the first 80% data as training data and the remaining 20% data as test data. We also calculate the training loss function using root mean squared error (RMSE) to measure how predictive our model concerning the training data.
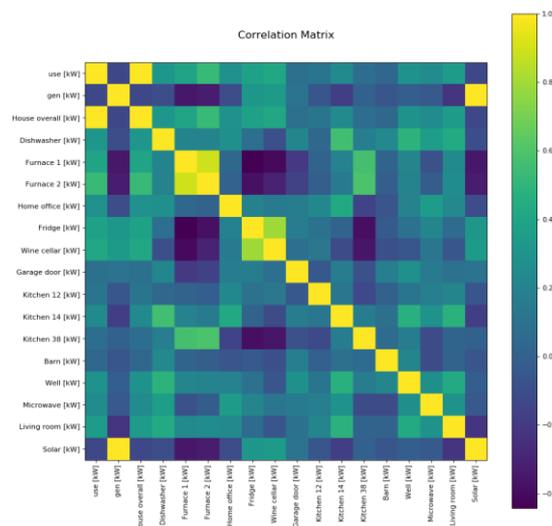
2.2. Dataset and exploratory analysis

In this section, we discuss the dataset and perform an exploratory analysis of the dataset.

2.2.1. Smart home dataset

For evaluating the performance of our forecasting system, we used IoT Smart Home data with weather information. This dataset is one week of historical electricity consumption data provided by Kaggle [13]. The data consist of 500,910 records of electricity and weather data that is recorded every second. There are 31 data features measured in this dataset: 18 electricity data features and 13 weather data features. We exclude weather dataset because there are too many data anomalies in the dataset, i.e. temperature that ranges from 12˚ to more than 90˚ in just seven days period. Data features and its description is shown in Table 1.

Table 1. Electricity data information

Data	Description
use [kW]	Total energy consumption
gen [kW]	Total energy generated by solar or other resources
House overall [kW]	Overall household energy consumption
Dishwasher [kW]	Energy consumed by the dishwasher
Furnace 1 [kW]	Energy consumed by furnace 1
Furnace 2 [kW]	Energy consumed by furnace 2
Home office [kW]	Energy consumed in home office
Fridge [kW]	Energy consumed by fridge
Wine cellar [kW]	Energy consumed by wine cellar
Garage door [kW]	Energy consumed by garage door
Kitchen 12 [kW]	Energy consumed in kitchen 1
Kitchen 14 [kW]	Energy consumed in kitchen 2
Kitchen 38 [kW]	Energy consumed in kitchen 3
Barn [kW]	Energy consumed by barn
Well [kW]	Energy consumed by well
Microwave [kW]	Energy consumed by microwave
Living room [kW]	Energy consumed in living room
Solar [kW]	Solar power generation

Fig 2. Correlation matrix for electricity consumption data

We check correlations between data features in electricity data using Pearson correlation method, which is a standard for Python Panda data frame correlation function [14] and the most widely used correlation method. This function finds the pairwise correlation of all columns or features in the dataset. It measures a linear association between features and calculates the correlation coefficient or degree of relationship between two features or variables. The correlation coefficient value is between -1 and 1. A value closer to 0

means a weaker correlation between the two features. A value closer to 1 or -1 implies stronger positive and negative correlation between the two features, respectively. Thus, the correlation of a variable or feature with itself will always return value 1.

The result is shown as a correlation matrix in Fig.2 . We can see that there are four features with strong correlation. `Use [kW],` and `House overall [kW]` are strongly correlated with `gen [kW]` and `Solar [kW]`, respectively. We exclude `use [kW]` for the experiment because it represents the same data as `House overall [kW]`, which records overall electricity consumption in the household. We also exclude `gen [kW]` because both `gen [kW]` and `Solar [kW]` represents the amount of energy generated by the solar panel.

Fig.3 shows the histogram for electricity consumption data from rooms and appliances in

Data	Description
use [kW]	Total energy consumption
gen [kW]	Total energy generated by solar or other resources
House overall [kW]	Overall household energy consumption
Dishwasher [kW]	Energy consumed by the dishwasher
Furnace 1 [kW]	Energy consumed by furnace 1
Furnace 2 [kW]	Energy consumed by furnace 2
Home office [kW]	Energy consumed in home office
Fridge [kW]	Energy consumed by fridge
Wine cellar [kW]	Energy consumed by wine cellar
Garage door [kW]	Energy consumed by garage door
Kitchen 12 [kW]	Energy consumed in kitchen 1
Kitchen 14 [kW]	Energy consumed in kitchen 2
Kitchen 38 [kW]	Energy consumed in kitchen 3
Barn [kW]	Energy consumed by barn
Well [kW]	Energy consumed by well
Microwave [kW]	Energy consumed by microwave
Living room [kW]	Energy consumed in living room
Solar [kW]	Solar power generation

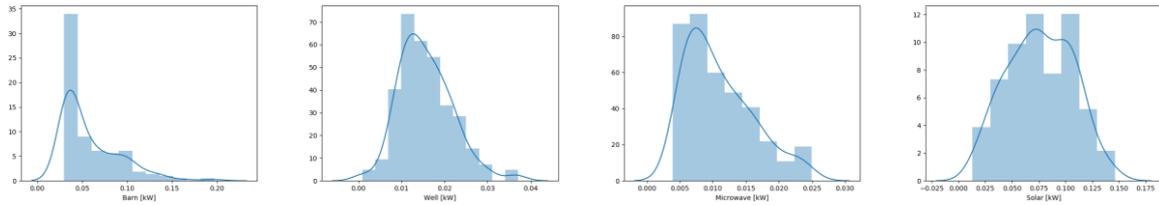. In the histogram, we can see the distribution for each electricity consumption data.
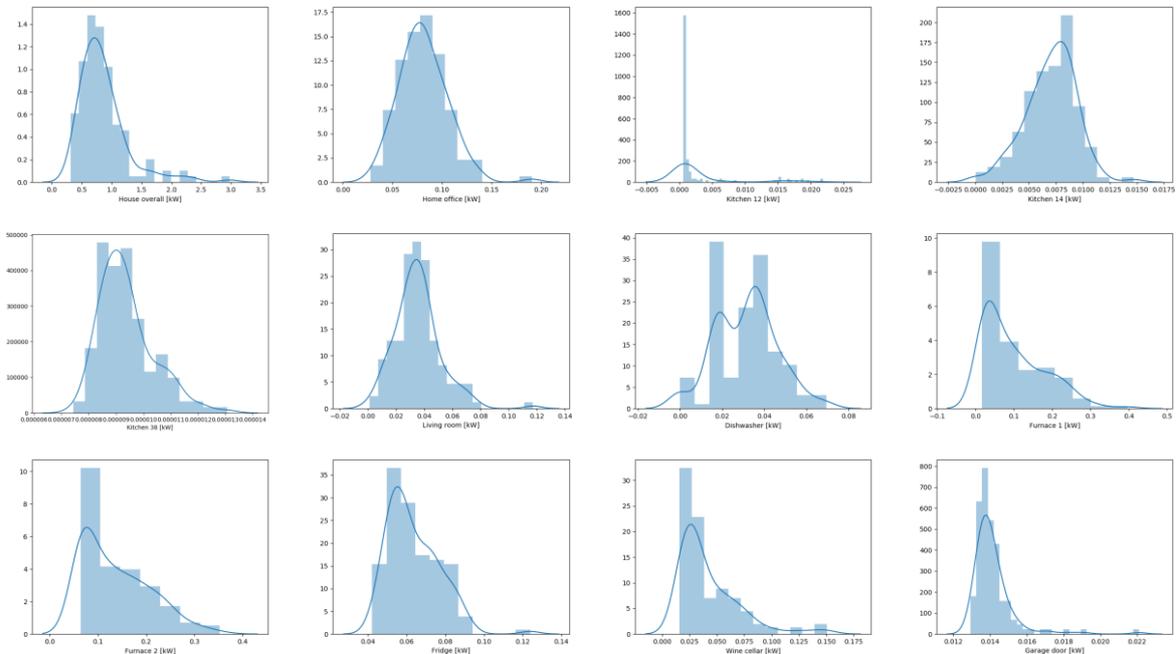
Fig 3. Histogram for room and appliances electricity consumption

2.2.2.　Time series features

Because of the high level of granularity in `time` feature, we applied to downsample to the dataset. First, we check feature importance of `time` data by creating `second`, `hour`, `minute`, `day`, and `day of week` feature from `time` feature. As shown in Fig. 4, an `hour` has the highest F score, followed by a `minute`, `day`, `day of week`, and `second`.

In the experiment, we consider using `hour` and `minute` to downsample the dataset. Since hour has the highest F score value compared to the other time feature and the limitation of data set size, we downsample the dataset into hourly data. Next, we also use the second highest time feature score, which is a `minute` time feature. We limit our experiment using `minute` time feature by 30 and 15 minutes because value lower than 15 implies higher granularity. The strategy to use hourly, 30 minutes and 15 minutes is also widely used in creating segmentation for household electricity consumption [10]. Therefore, we train the data model using three different downsampling time: 1 hour, 30 minutes, and 15 minutes.
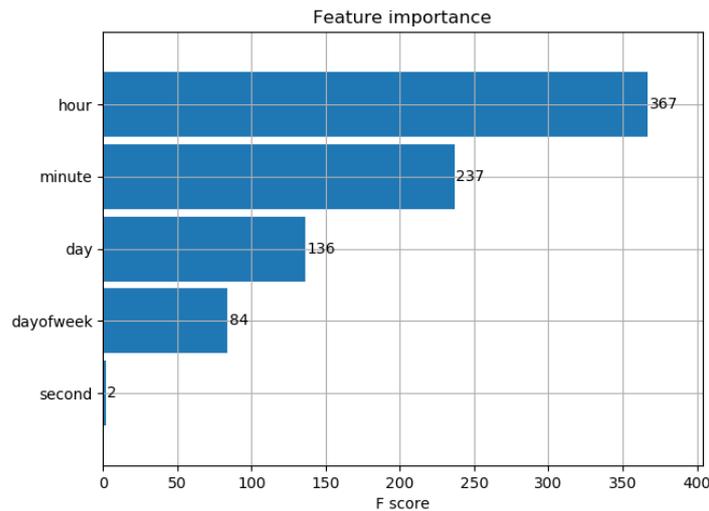

Fig 4. Feature importance

2.3.　Gradient boosting

There are only seven days worth of electricity consumption data in the dataset. This small data set size represents a challenge to train data for forecasting. We use Extreme Gradient Boosting algorithm to train our data model because it is widely recognized as a winning approach to solving a number of machine learning and data mining challenges [12] and it works well with a smaller number of data. Another forecasting algorithm, especially that utilizes moving average, such as ARIMA, usually needs a higher amount of data due to the nature of their sliding windows approach [4]-[6].

Extreme Gradient Boosting originally came from Gradient Boosting, which is used for supervised learning problem [15]. We train data x_i to predict a target variable y_i. Extreme Gradient Boosting uses decision tree ensembles, which consists of a set of classification and regression trees (CART) [12]. By using the ensemble model, it sums the prediction of multiples trees together and will give richer interpretation than the single tree. The objective function is denoted in (1):

$$obj(\theta) = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{k=i}^{K} \Omega(f_k) \tag{1}$$

where K is the number of trees, f is a functional space, and the objective function consists of training loss and regularization.

3.   RESULTS AND DISCUSSION

In this section, we report our experiments by evaluating the performance of the electricity consumption forecasting using an IoT Smart Home data set.

3.1.   Environment

We implemented the forecasting system in Python language using Python version 3.6.7. We used XGBoost library version 0.82 to train the dataset and use scikit-learn library version 0.20 for statistical computation. All experiments are conducted on a commodity machine with Intel Core i7-4500U 1.80 GHz CPU and 8 GB memory. This machine runs a Windows 7 64-bit operating system.
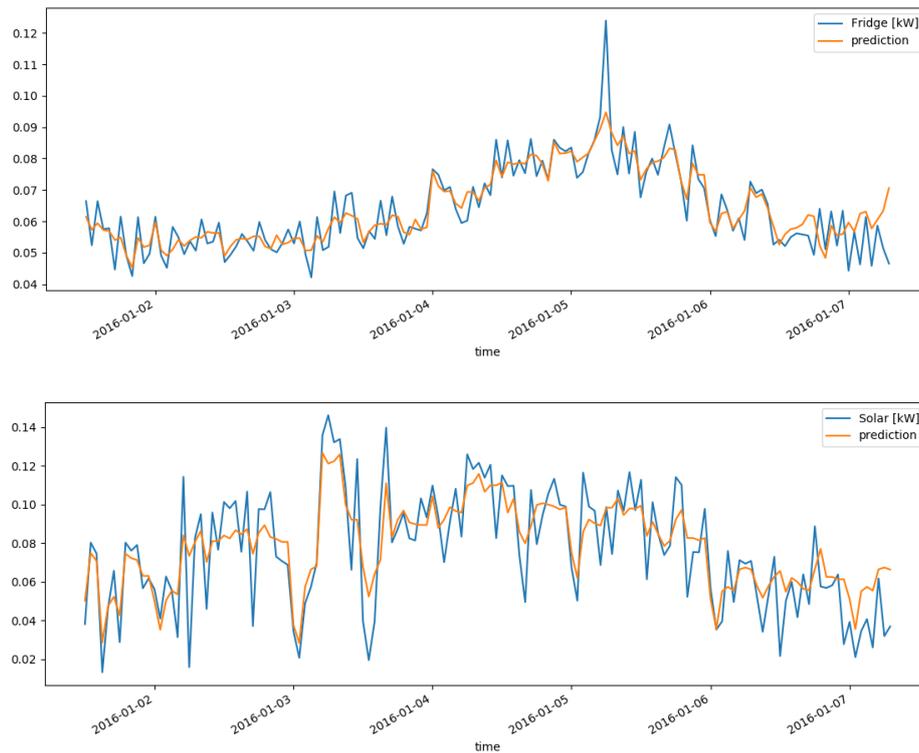
3.2.   Experiment result

Fig. 5 shows some of the forecasting results from room electricity consumption dataset. We can see that our forecasting system can make a good prediction using a small number of training data. Electricity consumption prediction in an overall house is worse than home office prediction, especially on January 5th where there is a sudden peak usage in electricity in the household.
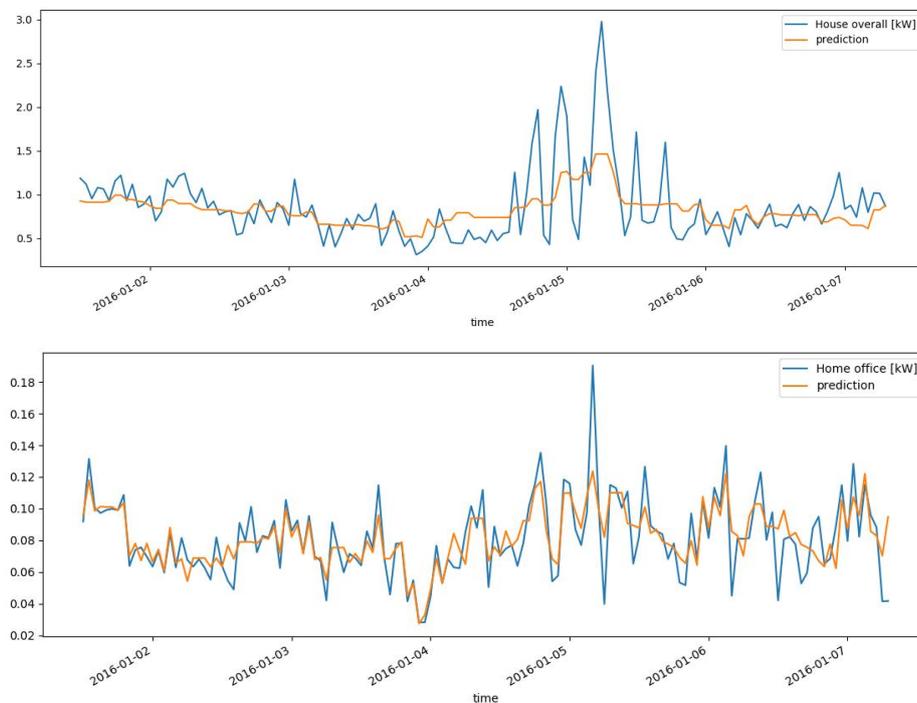
Fig 5. Actual and prediction data for room electricity consumption with 1 hour downsampling

Electricity consumption prediction for some individual appliances can be seen in Fig. 6. Prediction for fridge electricity consumption and solar power generation works quite well considering the size of training data.

Fig 6. Actual and prediction data for appliances with 1-hour downsampling

The overall experiment results are shown in Tabel 2. We calculate root mean square error (RMSE) for each electricity consumption prediction for each room and appliance in the household. As we can see, by using downsampling over 1-hour time period yields the best result compared to 15 minutes and 30 minutes downsampling time.

Another observation is that training error value is bigger than the test error value. This is usually a sign of underfitting. There are a few reasons for this to happen, one of them is because the training data is small. The easiest way to solve this issue is that we can add more data to adjust our prediction result.

Table 2. RMSE comparison for different downsampling

Data	RMSE (15 mins)		RMSE (30 mins)		RMSE (1 hour)	
	Test	Train	Test	Train	Test	Train
Room						
House overall [kW]	0.29695	0.46635	0.24938	0.36226	0.19904	0.32616
Home office [kW]	0.04231	0.03284	0.03347	0.02165	0.02201	0.01302
Kitchen 12 [kW]	0.01310	0.00882	0.01006	0.00603	0.00694	0.00401
Kitchen 14 [kW]	0.00428	0.00336	0.00325	0.00227	0.00254	0.00161
Living room [kW]	0.03256	0.02731	0.02341	0.02070	0.01743	0.01791
Appliances						
Dishwasher [kW]	0.03568	0.03453	0.02144	0.01925	0.00994	0.01157
Furnace 1 [kW]	0.06815	0.08705	0.06408	0.08399	0.05939	0.08341
Furnace 2 [kW]	0.05916	0.08745	0.05056	0.07289	0.04519	0.07207
Fridge [kW]	0.01416	0.01362	0.01089	0.00933	0.00955	0.00500
Wine cellar [kW]	0.01298	0.01329	0.01050	0.01178	0.01003	0.00978
Garage door [kW]	0.00143	0.00180	0.00098	0.00132	0.00054	0.00099

Barn [kW]	0.05523	0.04893	0.03918	0.03493	0.02938	0.02351
Well [kW]	0.01002	0.00988	0.00702	0.00713	0.00608	0.00446
Microwave [kW]	0.01191	0.00860	0.00967	0.00573	0.00674	0.00363
Solar [kW]	0.04458	0.05433	0.03118	0.02677	0.02066	0.01547

4.   CONCLUSION

In this paper, we have implemented the household electricity consumption forecasting system using extreme gradient boosting algorithm. This system can make a good prediction model by using the small size of training data. The best result is shown when we resample the data over 1-hour time period compared to 15 and 30 minutes. The RMSE score for training is still bigger than test data, one of the reason is that the size of the dataset. We can improve our predictive model by adding new data to address this issue.

REFERENCES

[1]   S. Sorrell, "Reducing energy demand: A review of issues, challenges and approaches," *Renewable and Sustainable Energy Reviews*, vol. 47, pp. 74–82, 2015, doi: 10.1016/j.rser.2015.03.002.

[2]   F. Karanfil and Y. Li, "Electricity consumption and economic growth: exploring panel-specific differences," *Energy Policy*, vol. 82, pp. 264–277, 2015, doi: 10.1016/j.enpol.2014.12.001.

[3]   S. Sarwar, W. Chen, and R. Waheed, "Electricity consumption, oil price and economic growth: Global perspective," *Renewable and Sustainable Energy Reviews*, vol. 76, pp. 9–18, 2017, doi: 10.1016/j.rser.2017.03.063.

[4]   Z. Mohamed and P. Bodger, "Forecasting Electricity Consumption A comparison of models for New Zealand," in *Electricity Engineers' Association of New Zealand Annual Conference*, 2004 doi: 10.1016/j.rser.2017.03.063.

[5]   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*, vol. 67, pp. 431–438, 2015, doi: 10.1016/j.ijepes.2014.12.036.

[6]   A. Veit, C. Goebel, R. Tidke, C. Doblander, and H.-A. Jacobsen, "Household Electricity Demand Forecasting-- Benchmarking State-of-the-Art Methods," in *Proceedings of the 5th International Conference on Future Energy Systems*, 2014, available at: https://arxiv.org/abs/1404.0200.

[7]   H. Guo, Q. Chen, Q. Xia, C. Kang, and X. Zhang, "A monthly electricity consumption forecasting method based on vector error correction model and self-adaptive screening method," *International Journal of Electrical Power and Energy Systems*, pp. 427–439, 2018, doi: 10.1016/j.ijepes.2017.09.011.

[8]   A. Hussain, M. Rahman, and J. A. Memon, "Forecasting electricity consumption in Pakistan: The way forward," *Energy Policy*, no. 90, pp. 73–80, 2016, doi: 10.1016/j.enpol.2015.11.028.

[9]   F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, 2015, doi: 10.1109/TSG.2014.2364233.

[10]   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, doi: 10.1109/TSG.2013.2278477.

[11]   S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE transactions on smart grid*, vol. 7, no. 1, pp. 136–144, 2016, doi: 10.1109/TSG.2015.2409786.

[12]   T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016, available at: https://arxiv.org/abs/1603.02754.

[13]   Kaggle, "Smart Home Dataset with Weather Information.", available at: Google.

[14]   Python, "Python Panda Dataframe.", available at: Google.

[15]   J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, pp. 1189– 1232, 2001, available at : Google Scholar.