A Review of the Main Machine Learning Methods for Predicting Residential Energy Consumption.

Alfonso González-Briones*†, Guillermo Hernández*, Tiago Pinto‡, Zita Vale‡, and Juan M. Corchado*†§¶
*BISITE Research Group, University of Salamanca. Edificio I+D+i, Calle Espejo 2, 37007. Salamanca, Spain.
†Air Institute, IoT Digital Innovation Hub, Carbajosa de la Sagrada, 37188. Salamanca, Spain.
‡GECAD Research Group, Polytechnic Institute of Porto, Porto, Portugal.
§Department of Electronics, Information and Communication, Faculty of Engineering,
Osaka Institute of Technology, 535-8585 Osaka, Japan.
¶Pusat Komputeran dan Informatik, Universiti Malaysia Kelantan, Karung Berkunci 36,
Pengkaan Chepa, 16100 Kota Bharu, Kelantan, Malaysia.

***Abstract*—The ability to predict future energy consumption is very important for energy distribution companies because it allows them to estimate energy needs and supply them accordingly. Consumption prediction makes it possible for those companies to optimize their processes by, for example, providing them with knowledge about future periods of high energy demand or by enabling them to adapt their tariffs to customer consumption. Machine Learning techniques allow to predict future energy consumption on the basis of the customers' historical consumption and several other parameters. This article reviews some of the main machine learning models capable of predicting energy consumption, in our case study we use a specific set of data extracted from a two-year-period of a shoe store. Among the evaluated methods, Gradient Boosting has obtained an 86.3% success rate in predicting consumption.**

***Index Terms*—Energy Forecasting, Machine Learning, Gradient Boosting, XGBoost, Lasso, Ridge regression, SGDRegressor, MLP**

I. Introduction

Machine Learning is a subfield of Artificial Intelligence, it focuses on the development of systems with an automated learning capacity. This learning can be used to identify complex patterns through the analysis of large amounts of data. Machine Learning algorithms are learning models that process data by creating an identification, classification or prediction mechanism, either supervised or unsupervised.

Machine Learning algorithms provide greater insight into the information contained in large amounts of diverse data, making it possible for much more knowledgeable and precise decision-making than would ever be possible with manual data analyses [1], [2]. These algorithms allow us to detect behavioral patterns in a given data set and to identify the key variables that affect trends and cause changes in the pattern of the data [3], [4].

These capabilities have made Machine Learning models applicable to the field of energy, where they can serve for multiple purposes. One of their uses is the prediction of the energy consumers will consume in the future, so that companies can adapt their tariffs to energy consumption or manage the energy supply according to demand, preventing power outages. To be able to predict consumption it is necessary to have the customers' historical data, which are used in their preprocessed form to train the models. This article reviews some machine learning algorithms designed for predicting energy consumption, in the case study the models are trained with a historical data set from the past year and the best performing algorithm among those studied is identified.

This article is organized as follows: section 2 describes the state of the art of machine learning models used in energy prediction, Section 3 describes the proposal, and Section 4 presents the results and conclusions.

II. Machine Learning Model Review for Energy Consumption Prediction

In this section we review the different machine learning methods that allow us to predict energy consumption by means of the historical data sets of the customers' energy consumption. In order to review the methods that allow us to achieve the goal of knowing which model produces the best energy prediction results, we will review the most commonly used methods in the literature.

This article reviews the most popularly employed machine learning models found in the literature. Specifically, the algorithms designed to predict energy consumption, helping identify among the studied models the one that gives the best results, to this end, we trained the algorithms with data from a shoe store.

One of the first techniques used to predict energy consumption were artificial neural networks (ANN), [5] and [6]. Neural networks are a simple model that can be used in this field thanks to the advances in calculation capacity and the use of GPUs, minimizing a typical problem experienced by this model; the lengthy training time. This problem was caused by the size of the data set which must be quite large for the NN to achieve noticeable results. Hippert et al. conducted a review of the use of NN in energy consumption prediction problems, their study identifies new reasons for which researchers discard this methodology as a means of predicting consumption [7]. Often the results, although apparently good, are not convincing, as errors are not evaluated correctly. In addition, more recent studies show that results are better with other models such as Linear Regression or Support Vector Regression [7], [8]. In this study, a review of the MLP (Multi-Layer Perceptron) model will be performed.

Gradient Boosting is an automatic learning technique used in regression analysis and statistical classification problems, which produces a predictive model in the form of a set of weak predictive models, typically decision trees. It builds the model in a tiered manner as other boosting methods do, and generalizes them allowing for the arbitrary optimization of a differentiable loss function.

Gradient Boosting is a model that has been employed in multiple papers on energy consumption prediction, such as the following: [9], [10] and [11]. In these works it is shown how Gradient Boosting is an effective prediction algorithm for both classification and regression tasks. By selecting the number of components included in the model, the so-called compensation of bias-variance can be controlled easily in the estimation.

A study similar to ours but focused only on the use of Gradient Boosting, has been conducted by Souhaib Ben Taieb et al. in [12], who, with the aim of predicting energy consumption, collected data from Smart meters with information on the users' energy habits. The authors propose to estimate an additive regression model by quintiles for a set of quintiles of the future distribution by means of a boosting procedure. In this work they used data collected from 3639 households at 30-minute intervals over a year and a half.

The empirical results demonstrate that the quantum regression approach provides more precise predictions for disaggregated demand, whereas the traditional normal assumption approach is a better approximation for aggregate demand. These results are particularly useful, as more disaggregated energy data will become available in the future.

Extreme Gradient Boosting is based on the principles of gradient reinforcement structure and is designed to "push the calculation limits of machines to provide a scalable, portable and accurate library". XGBoost is one of the implementations of the Gradient Boosting concept, but what makes XGBoost unique is that it uses "a more regularized formalization model to control overtuning, which gives it better performance. This algorithm has been used in numerous articles to predict, for example, crude oil price or public transport demand. In the field of energy prediction it has been used in researches such as [13] and [14]. The paper of Li and Zhang [15] uses a hybrid approach to predicting the security of the energy supply based on ARIMA and XGBoost. This hybrid approach is much more accurate than other models the authors had evaluated using XGBoost. The results produced by the use of XGBoost have been satisfactory in most of the state-of-the-art literature, for this reason we have included this method in our evaluation.

Least absolute shrinkage and selection operator (Lasso) is a regression analysis method that performs variable selection and regularization to improve the accuracy and interpretability of the statistical model produced by it [16]. Lasso improves the accuracy of predictions and the interpretability of statistical regression models by altering the model construction process by selecting only a subset of the variables provided for use in the final model. Lasso was originally formulated for least squares models and this simple case reveals a substantial amount about the behavior of the estimator, including its relationship to Ridge regression and best subset selection and the connections between lasso coefficient estimates and the so-called soft thresholding[17]. It also reveals that (like standard linear regression) the coefficient estimates do not need to be unique if covariates are collinear. Lasso is able to achieve both of these goals by forcing the sum of the absolute value of the regression coefficients to be less than a fixed value, which forces certain coefficients to be set to zero, effectively choosing a simpler model that does not include those coefficients. This idea is similar to Ridge regression, in which the sum of the squares of the coefficients is forced to be less than a fixed value, though in the case of Ridge regression, this only shrinks the size of the coefficients, it does not set any of them to zero. Ridge regression is the most commonly used method of regularization of ill-posed problems. Ridge regression is a similar technique to LASSO, but it uses a quadratic penalty term where LASSO uses a linear one. Generally speaking, ridge regression performs better when a small amount of regularization is required for a large number of predictor variables, while LASSO performs better when a large amount of regularization is required for a small number of predictor variables. Elastic net is another algorithm that combines the features of both. Ridge regression has been used for travel demand prediction [18], tourist demand [19] or wind speed prediction [20]. This model has also been studied in the field of energy consumption prediction, for example in [21] and [22].

Stochastic Gradient Descent Regressor (SGDRegressor) is an iterative method for optimizing an objective function with suitable smoothness properties. It is called stochastic because the method uses randomly selected (or shuffled) samples to evaluate the gradients, hence SGD can be regarded as a stochastic approximation of gradient descent optimization. SGDRegressor is a simple yet very efficient approach to discriminative learning of linear classifiers under convex loss functions such as (linear) Support Vector Machines and Logistic Regression. SGD has been successfully applied to large-scale and sparse machine learning problems often encountered in text classification and natural language processing. Given that the data is sparse, the classifiers in this module easily scale to problems with more than 10^5 training examples and more than 10^5 features. Examples of how this method is applied to the prediction of energy consumption can be found in the following works [23] and [24].

From the present review of the state of the art, it is clear which machine learning models have the best rate of success in varied data sets. The proposed system will allow us to assess which of these models provide predictions with a higher precision rate for a shoe store data set, a system will be developed that will implement the following models MLP, Gradient Boosting, XGBoost, Lasso, Ridge regression and SGDRegressor.

III. Study of the Data Set for the Application of Prediction Models

Before starting with the application of the machine learning models included in the review, it is necessary to prepare the set of data to which they are going to be applied. This process is key to achieving a satisfactory success rate. For the evaluation of machine learning models, the system has used a set of consumption data from a shoe store located in Salamanca, Spain. The data in the data set belong to the following time interval between 05/01/2016 and 11/12/2018. The data set consists of the date: day, month and year, distinguishing between weekdays and weekends, the electricity consumed (kWh) the previous day and current electricity consumption (kWh).

The data set has been modified to also include the electricity consumed in the previous day. This makes it possible to improve the prediction capacity of the algorithms. Although there is a natural relationship between the day of the year and the energy consumption, the strong variations in the latter due to external causes make this an insufficient predictor, as evidenced by the low values of the Pearson correlation index. Figure 1. shows the Pearson correlation that reveals the importance of including the variable Previous day, with r = 0.921. To complete this information, the consumption of the previous day has been used as an additional attribute, with which there is a clear correlation, as shown in the right part of the same figure.

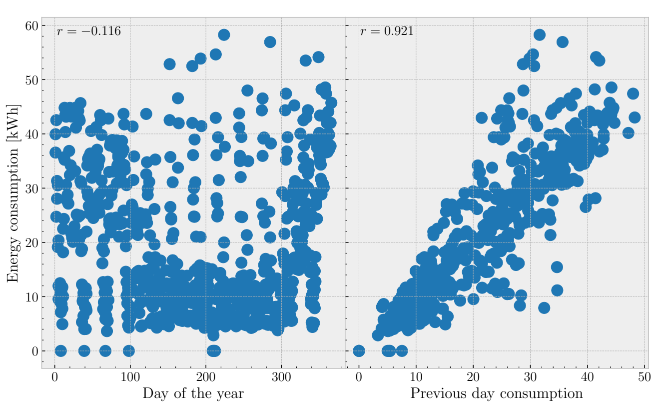

Fig. 1. Scatter plots of the day of the year (left) and the previous day energy consumption (right) with the actual energy consumption (vertical axis). The Pearson correlation coefficients are also shown in the figure.

Weekends are also important periods in which energy consumption must be predicted, as shown by the conditional distributions in Figure 2.

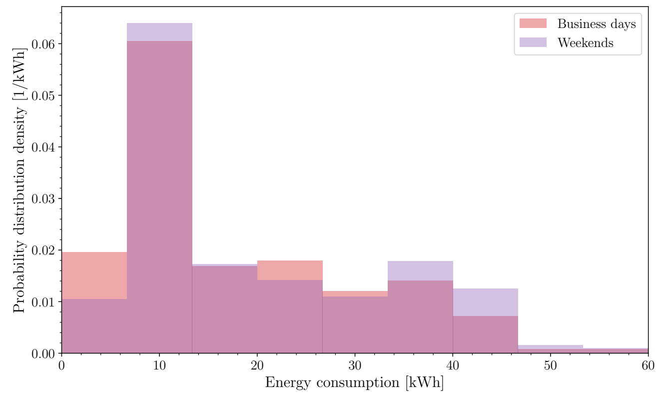

Fig. 2. Conditional distributions of energy consumption between week days and weekends.

To provide a better representation of the day of the year, it has been linearly scaled in the range [0,1], continuously mapping the summer solstice to 1 and winter solstice to 0.

IV. Results

Once a data set has been prepared, the prediction process begins. To train the models it is necessary to establish the best possible conditions, enabling in this way a successful prediction processes. Different methods have been employed in building the said Models, using the transformed day of the year, the previous day energy consumption, and the business day or weekend condition. Actual vs predicted values are shown in Figure 3.

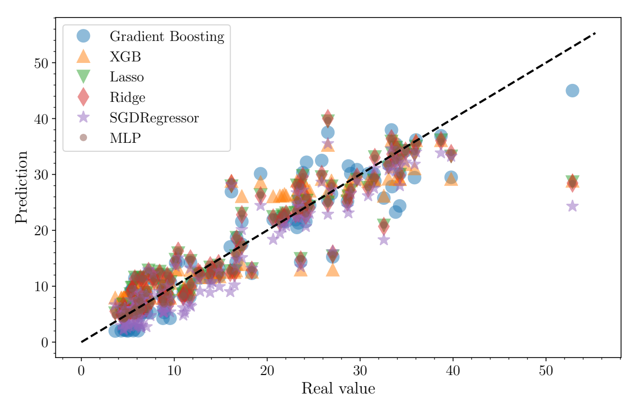

Fig. 3. Scatter plots of the actual values vs predicted values of every machine learning model used by the system.

TABLE I., shows quite similar prediction results in which Gradient Boosting stands out. However, these results can be improved by using a selection of parameters that are used to enhance the training process.

TABLE I
MACHINE LEARNING METHODS SCORE WITHOUT PARAMETER SELECTION.

Method	Accuracy
Gradient Boosting	0.862
XGB	0.824
Lasso	0.857
Ridge	0.857
SGDRegressor	0.754
MLP	0.856

The individual results of each Machine Learning model are shown below using a selection of the parameters that produce the best prediction results. In Figure 4 we can see the results of Gradient Boosting method.

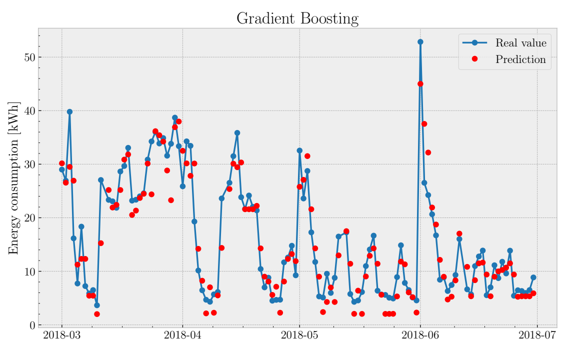

Fig. 4. Scatter plots of the prediction results of the Gradient Boosting model.

In Figure 5. we can see the results of XGB.

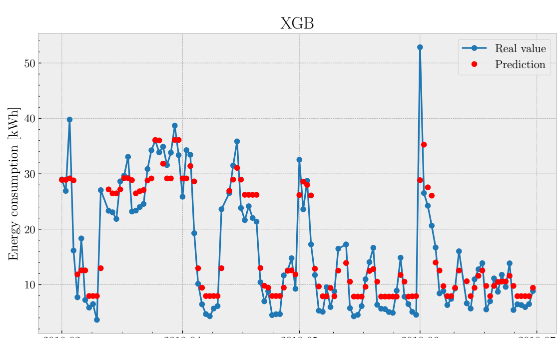

Fig. 5. Scatter plots of the prediction results of XGB.

In Figure 6 we can see the results of Lasso.

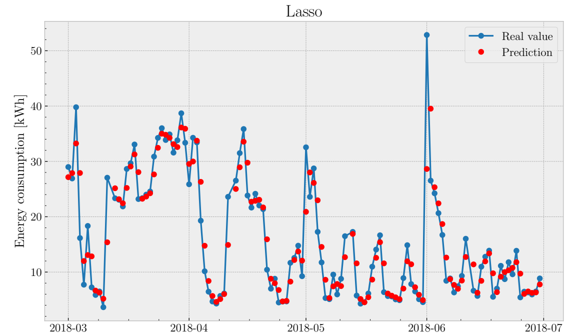

Fig. 6. Scatter plots of the prediction results of Lasso.

In Figure 7 we can see the results of Ridge regression.

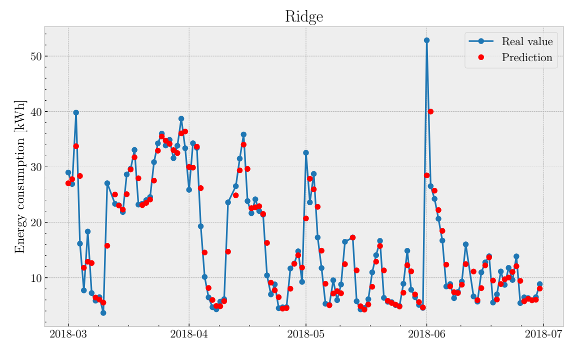

Fig. 7. Scatter plots of the prediction results of Ridge regression model.

In Figure 8 we can see the results of SGDRegressor.

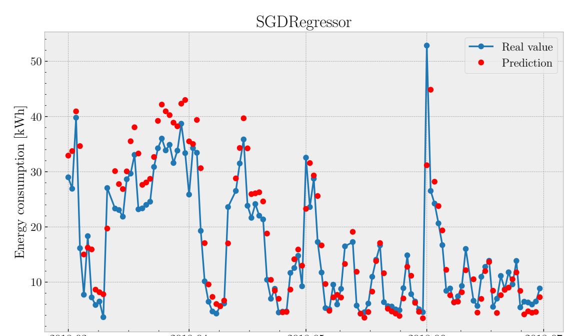

Fig. 8. Scatter plots of the prediction results of SGDRegressor model.

In Figure 9 we can see the results of MLP.

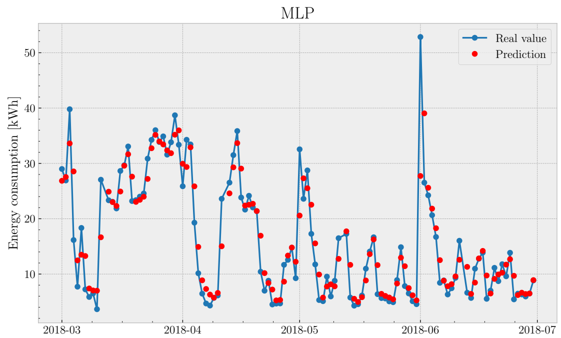

Fig. 9. Scatter plots of the prediction results of MLP.

TABLE II shows the level of precision of the machine learning models used, using the method of selecting the parameters that produce the best results in the prediction process for each of the models. It can be seen how there is a slight improvement over the method that uses all the parameters in the prediction.

TABLE II
MACHINE LEARNING METHODS' SCORE WHEN TRAINED WITH A SELECTION OF THE BEST PARAMETERS.

Method	**Accuracy**
Gradient Boosting	0.863
XGB	0.824
Lasso	0.857
Ridge	0.857
SGDRegressor	0.801
MLP	0.856

V. CONCLUSIONS AND FUTURE WORK

In this work some of the main machine learning models have been considered. Our review was focused specifically on ML models designed for the prediction of energy consumption. The machine learning methods used in this work are some of the models with the best prediction results according to the literature (Gradient Boosting, XGBoost, Lasso, Ridge regression, SGDRegressor, MLP). These models have been applied to a two-year-period data set containing information from a shoe store.

These data have made it possible to find out that Gradient Boosting has achieved the highest level of accuracy of 86.3% , out of all the models studied. Lasso, Ridge regression obtained 85.7%, MLP 85.6%, XGBoost 82.4% and finally SGDRegressor 80.1% being the model with the worst results. This comparison does not mean that Gradient Boosting is better than the rest of the models, its use is simply more suitable to the variables that make up the data set (day, day of the week, week, presence, etc.).

ACKNOWLEDGMENT

This work was carried out under the frame of the "Towards Sustainable Intelligent Mobility: Blockchain-based framework for IoT Security" Ref. RTI2018-095390-B-C32" project. The project was supported and funded by the Spanish Ministerio de Economía, Industria y Competitividad. Retos de investigación, Proyectos I+D+i.

REFERENCES

[1] A. González-Briones, J. A. Castellanos-Garzón, Y. Mezquita-Martín, J. Prieto, and J. M. Corchado, "A multi-agent system framework for autonomous crop irrigation," in *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*. IEEE, 2019, pp. 1–6.

[2] A. Rivas, J. M. Fraile, P. Chamoso, A. González-Briones, I. Sittón, and J. M. Corchado, "A predictive maintenance model using recurrent neural networks," in *International Workshop on Soft Computing Models in Industrial and Environmental Applications*. Springer, 2019, pp. 261–270.

[3] A. J. Sánchez, S. Rodríguez, F. de la Prieta, and A. González, "Adaptive interface ecosystems in smart cities control systems," *Future Generation Computer Systems*, vol. 101, pp. 605–620, 2019.

[4] A. González-Briones, P. Chamoso, R. Casado-Vara, A. Rivas, S. Omatu, and J. M. Corchado, "Internet of things platform to encourage recycling in a smart city," 2019.

[5] A. Ahmad, M. Hassan, M. Abdullah, H. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ann and svm for building electrical energy consumption forecasting," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 102–109, 2014.

[6] C. Hamzaçebi, "Forecasting of turkey's net electricity energy consumption on sectoral bases," *Energy policy*, vol. 35, no. 3, pp. 2009–2016, 2007.

[7] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on power systems*, vol. 16, no. 1, pp. 44–55, 2001.

[8] H.-x. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012.

[9] S. B. Taieb and R. J. Hyndman, "A gradient boosting approach to the kaggle load forecasting competition," *International journal of forecasting*, vol. 30, no. 2, pp. 382–394, 2014.

[10] C. Persson, P. Bacher, T. Shiga, and H. Madsen, "Multi-site solar power forecasting using gradient boosted regression trees," *Solar Energy*, vol. 150, pp. 423–436, 2017.

[11] M. Landry, T. P. Erlinger, D. Patschke, and C. Varrichio, "Probabilistic gradient boosting machines for gefcom2014 wind forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1061–1066, 2016.

[12] S. B. Taieb, R. Huser, R. J. Hyndman, and M. G. Genton, "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2448–2455, 2016.

[13] H. Zheng, J. Yuan, and L. Chen, "Short-term load forecasting using emd-lstm neural networks with a xgboost algorithm for feature importance evaluation," *Energies*, vol. 10, no. 8, p. 1168, 2017.

[14] W. Zhang, H. Quan, and D. Srinivasan, "Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination," *Energy*, vol. 160, pp. 810–819, 2018.

[15] P. Li and J.-S. Zhang, "A new hybrid method for china's energy supply security forecasting based on arima and xgboost," *Energies*, vol. 11, no. 7, p. 1687, 2018.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[17] V. Roth, "The generalized lasso," *IEEE transactions on neural networks*, vol. 15, no. 1, pp. 16–28, 2004.

[18] E. T. Fujii and J. Mak, "Forecasting travel demand when the explanatory variables are highly correlated," *Journal of Travel Research*, vol. 18, no. 4, pp. 31–34, 1980.

[19] E. Fujii and J. Mak, "Forecasting tourism demand; some methodological issues," *The Annals of Regional Science*, vol. 15, no. 2, pp. 72–82, 1981.

[20] F. Douak, F. Melgani, and N. Benoudjit, "Kernel ridge regression with active learning for wind speed prediction," *Applied energy*, vol. 103, pp. 328–340, 2013.

[21] N. Fumo and M. R. Biswas, “Regression analysis for prediction of residential energy consumption,” *Renewable and sustainable energy reviews*, vol. 47, pp. 332–343, 2015.
[22] P. Mirowski, S. Chen, T. K. Ho, and C.-N. Yu, “Demand forecasting in smart grids,” *Bell Labs technical journal*, vol. 18, no. 4, pp. 135–158, 2014.
[23] F. Rodrigues, C. Cardeira, and J. M. F. Calado, “The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in portugal,” *Energy Procedia*, vol. 62, pp. 220–229, 2014.
[24] L. Ekonomou, “Greek long-term energy consumption prediction using artificial neural networks,” *Energy*, vol. 35, no. 2, pp. 512–517, 2010.