Proc. of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME 2023)
19-21 July 2023, Tenerife, Canary Islands, Spain

2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) | 979-8-3503-2297-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICECCME57830.2023.10252919

Development of surplus power generation forecast for use by residential loads

Paloma G. S. Dias
Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança,
Campus de Santa Apolónia, 5300-253 Bragança, Portugal
Laboratório para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança
Campus de Santa Apolónia, 5300-253 Bragança, Portugal
palomapaes@live.com

Thadeu Brito
CeDRI, SusTEC
Faculty of Engineering of University of Porto (FEUP), Porto, Portugal
brito@ipb.pt

William R. Silva
CeDRI, SusTEC
a52488@alunos.ipb.pt

Ana I. Pereira
CeDRI, SusTEC
apereira@ipb.pt

Luis C. G. Lopes
Departament of Electroelectronics (DEELP),
Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Leopoldina-MG, Brazil
lcgamboa@cefetmg.br

Murillo F. dos Santos
DEELP, CEFET-MG
murillo.ferreira@cefetmg.br

Paulo Costa
FEUP
paco@fe.up.pt

José Lima
CeDRI, SusTEC
jllima@ipb.pt

***Abstract*—Energy consumption has been increasing in the last years and thus, energy efficiency is one of the most important topics actually. Besides, the consumption and energy generation forecast help in efficiency optimization. This paper presents the development of a system for forecasting surplus power generation to be used by residential loads connected to smart plugs. In this way, it is intended to collaborate with the use of surplus energy production in electrical devices in a residence instead of sending to batteries or to the grid. This work presents the theoretical basis of the project and the architecture of the developed system. A Machine Learning method applied to photovoltaic generation data in a residence was used to predict surplus energy.**

***Index Terms*—Surplus energy, Data forecasting, Machine Learning, Internet of Things**

I. INTRODUCTION

Government incentives for domestic energy production can take various forms, such as tax credits, subsidies, and regulatory measures. In recent years, several governments have enforced their policies, and the growth of people producing their own power using renewable resources is noticed worldwide. However, some microgrids can be designed to make more power than a daily household can utilize. To handle the surplus of power production, devices/applications can be used to manage their own usage and sell the extra to a central grid. However, considering the high costs of purchasing a battery bank and ecological issues, the project may be unattainable.

Based on that, smart energy management systems have become a focus of technology development in the energy sector. Some implementations of these systems include the Internet of Things (IoT) to ensure the connection between parts of the communication network and secure data transport [1]. In this sense, consumers can remotely monitor and control household devices, allowing for energy savings and conscious electricity consumption. But sometimes, even remotely, only turning on/off the equipments is not enough to save energy. On the other hand, controlling the load of a specific device could save power according to the production [2], [3]. Additionally, Home Energy Management Systems (HEMS) applications can be connected to microgrids or main grids, reducing the use of fossil fuels [4].

One of the differentials of the smart plugs (considered in this work) is the ability to control not only the appliances' on/off but also operating power based on the load to which they are connected. With this in mind, the objective of the proposed work is to develop a monitoring and control system using Machine Learning (ML) to manage the energy consumption by residential loads connected to smart plugs. For this control to be possible, the system must have access to the user's excess generation and consumption information, which is provided by the plugs. Based on this, the system will be able to predict these variables and automatically make decisions about directing excess energy to specific loads in the home.

In order to avoid costs in the installations, the final decision must use the IoT to control the devices. But the system must not be limited by this automation based on ML, this control can also be done manually by the users. For this reason, the focus of this work does not concern how to perform this control but rather the prediction strategies used for automatic decision-making. Thus, the main contributions of this work are defined by:

- Predict surplus residential power generation to be used by a variable load.
- Analyze the efficiency of the prediction models with different data volumes and features.
- Based on the forecasting capability of the system, incentive the use of surplus energy to power residential loads instead of sending it to the grid or batteries.

The remainder of the paper is structured as follows: after this introduction section, Section II will present related works using ML to predict energy surplus. Section III will demonstrate the system architecture, whereas the next section (Section IV) will focus on the algorithm parameters. Section V presents the results, and the last section points out the conclusions and future work (Section VI).

II. RELATED WORKS

The objective is to predict energy data using ML, then the concept involved must be considered. Using ML, a system can learn from data or experiments to make decisions based on a model [5], [6]. Analyzing some case studies of its use in the context of energy management, [7] presents the use of submeters to measure the voltage and current of loads separately, considering that a Non-Intrusive Load Monitoring (NILM) system checks for separate consumption of loads. Based on these measurements, the authors aimed to predict from ML using Matlab as a tool. Among the methods applied are Nearest Neighbor Algorithm and Markov Chain.

In this work, the ML method adopted to predict surplus power generation is Linear Regression. Works like [8] use this method, in which the authors applied Multiple Linear Regression to perform the prediction of electricity consumption in commercial buildings. Seventeen predictor attributes were used for the predictive analysis of Y, the variable to be predicted which, in this case, corresponds to the annual energy consumption, in 7 building shapes.

Considering the vast application of ML for data prediction and decision-making in the context of residential energy management, it is noticeable the importance of developing smart systems in this sphere, ensuring the safety, reliability, economy, and comfort of consumers. In this sense, the related works demonstrate solutions with high rates of ACC to predict the surplus of energy production. But none of them aims to define a forecast with the smallest possible volume of data. Therefore, this work is dedicated to identifying the smallest possible interval to apply Linear Regression in an open-source dataset of residential loads and energy production.

III. SYSTEM ARCHITECTURE

In detail, the problem approached in this work is the lack of possibility of using the surplus energy generated by residential loads. On this, developing a predictive analysis of surplus generation for the future feeding of an automatic decision-making system able to define the loads that will use the surplus produced becomes a possible solution. The focus, then, is the implementation of an ML method for the prediction of surplus generation, based on generation data from a residence.

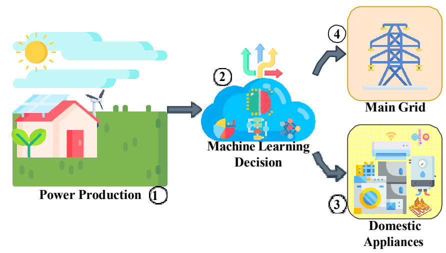


Fig. 1. Architecture of the proposed system.

The architecture of the proposed system is exemplified in Fig. 1, where there are four main topics. In **1**, there is the domestic production of electricity, generally, this type of production is based on renewable sources. Considering these cases of energy production, some variations in the power produced during the day are expected, for example, in the production through photovoltaic panels. Therefore, item **2**, which is the main objective of this work, is applied to the function of the intelligent system to predict surplus energy generation according to the user's production and consumption data. Through ML analysis, the system will decide the best destination for the extra power, whether it should go to item **3** (domestic appliances) or item **4** (sell to the main grid).

The prediction of the energy surplus produced indicated by an ML algorithm needs some evaluation metrics to be defined, as well as a database for learning as detailed ahead.

A. Metrics definitions

In general, ML uses data to perform decision-making by means of computational algorithms [9]. It is possible to highlight Linear Regression as one of the supervised learning methods of ML [10] defined by the relation presented in (1).

$$Y = a_0 X + a_1 + e \quad (1)$$

where Y represents the variable to be predicted, X is the predictor value, a_0 is the angular coefficient, a_1 is the linear coefficient, and e is the random error.

To apply this method to a database of surplus energy, tools such as the Python programming language, its libraries `Pandas` and `Scikit-Learn`, were used, and Visual Studio Code as a programming environment. After applying this method and obtaining the predicted values, important parameters in the predictive analysis must be considered: ACC and Mean Absolute Error (MAE). The first one measures the ACC of the model in predicting the correct result, and the second one indicates the MAE between the predicted and the actual values. The MAE represents the ratio between the sum of the error modules and the number of samples that would be predicted, considering the error as the difference between the actual and the predicted value.In addition, Median Absolute Error (MdAE), represents the median between predicted and actual values. Explained Variance Score (EVS) is another important parameter that represents the ability of the model to explain the variation in the data. Furthermore, Error Variance (E_{Var}) is the measure of the dispersion of the forecast errors relative to the mean of the errors. Another parameter considered among the metrics obtained in the analysis of the forecast models is the coefficient of determination, or R^2, which represents how effectively the independent variable X is able to explain the variation of the dependent variable Y. It is a metric with a scale from 0 to 1, and the closer to 1, the better X explains the variation of Y.

B. Database for learning

Regarding the database of surplus power generation, the `Package Household Data` dataset available on [11] was used, specifically the Excel file `household_data.xlsx`. It has consumption data and photovoltaic generation in kWh in industries and households in southern Germany over 4 years, between December 11th, 2014, at 5:45 pm to May 1st, 2019, at 10 pm, with 15-minute intervals of sample rate. We chose the consumption unit defined in the spreadsheet as *DE_KN_residential6* and the 1-year period to be studied: May 14th, 2016 to May 14th, 2017. It is noted that this residence located in an urban area has the following parameters:

- *DE_KN_residential6_circulationpump* represents a numerical quantity of the float type that shows the energy consumption in kWh of a circulation pump;
- *DE_KN_residential6_dishwasher* represents the energy consumption in kWh of a dishwasher;
- *DE_KN_residential6_freezer* represents the energy consumption in kWh of a freezer;
- *DE_KN_residential6_washingmachine* is the energy consumption in kWh of a washing machine;
- *DE_KN_residential6_grid_import* is the energy in kWh imported from the grid;
- *DE_KN_residential6_grid_export* is the energy in kWh exported to the main grid;
- *DE_KN_residential6_pv* is the total photovoltaic energy generated at a residential building in kWh.

All these values presented and made available in the dataset are cumulative. In this sense, it was necessary to calculate the difference between a cumulative sample and the previous one. Furthermore, after data normalization, the surplus energy data was defined.

IV. ALGORITHM PARAMETERS

For the application of Linear Regression, it must be considered that only energy values were analyzed, which does not include time variables. It was necessary to divide the X and Y values into test and training data, respectively. Thus, 80 % of the X and Y values were defined as training data, and 20 % as test data, for the creation of the Linear Regression models. Then, the predict function can use the test data to make predictions based on the model created from the training data. In the data analysis, in [12], [13], the authors considered the values of X or forecasters as the previous days, understanding that their data were daily. Thus, knowing that the data in this work are analyzed in 15-minute intervals, X corresponds to the previous measurements. In this way, understanding that Y are the values that should be predicted, $X1$ corresponds to the actual values read from Y considering 15 minutes earlier, $X2$ considers 30 minutes earlier, $X3$ considers 45 minutes earlier, $X4$ considers 60 minutes earlier, $X5$ considers 75 minutes earlier and $X6$ considers 75 minutes earlier. In this work, the inclusion of 6 features, from $X1$ to $X6$, is adopted.

After defining the Linear Regression model, it is important to acquire its coefficients to define the line function, according to (1), without random error, and insert the actual Y data considering previous measures in X. With the coefficients and the values X, which are the actual values of Y considering previous measurements, it is possible to find the predicted values Y. Another important detail is the implementation of not predicting negative values, a point that was treated based on the condition that if Y is less than zero, it becomes zero. In this way, considering that there are actual values of surplus energy that are negative, for the case where generation is less than consumption, the developed models will adopt the null value as the prediction result. Finally, the parameters mentioned in Section III were calculated, such as ACC, R^2, MAE, MdAE, E_{Var}, and EVS.

The next step was the predictive analysis, considering different periods of time and features for X ($X1$ to $X6$) to analyze the models that presented higher forecasting efficiency. The analyses performed were forecast with hours, forecasts from one day to the next, using 5 days for forecast, using 10 days for forecast and using 30 days for forecast.

V. RESULTS

The results obtained are based on the application of Linear Regression as a supervised ML method. There is no activation function and initialization method. For each data set, 6 models are developed, based on 1 to 6 features for X. Some parameter analyses will be considered to understand the efficiency of the models.

A. Forecast with hours

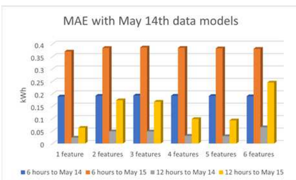
The first case involves using 6 and 12 hours on May 14th, 21st and 28th, and June 4th, to predict the rest of the day of May 15th, 22nd and 29th, and June 5th, respectively. Fig. 2 presents the MAE values obtained, considering the models using up to 6 new features for X.

In Fig. 2(a), the 12-hour model using 1 feature for X showed the lowest MAE, especially predicting May 14th. Furthermore, in 2(b), the 12-hour model with 1 feature for X shows the lowest MAE for both the forecast for the remainder of May 21st and the forecast for May 22nd. Models with 2 to 4

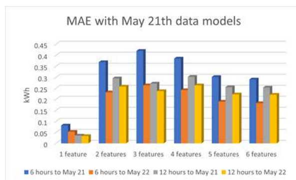
TABLE I. R^2 for May 14^{th}, 21^{st} and 28^{th}, and June 4^{th} from 6-hour and 12-hour models

Features	May 14^{th}		May 21^{st}		May 28^{th}		Jun 4^{th}	
	R^2 for 6 hours	R^2 for 12 hours	R^2 for 6 hours	R^2 for 12 hours	R^2 for 6 hours	R^2 for 12 hours	R^2 for 6 hours	R^2 for 12 hours
1	0.701	0.851	0.963	0.995	0.809	0.476	0.940	0.796
2	0.740	0.853	0.971	0.997	0.814	0.506	0.950	0.827
3	0.740	0.862	0.971	0.997	0.954	0.513	0.951	0.862
4	0.741	0.877	0.971	0.997	0.980	0.528	0.953	0.870
5	0.742	0.877	0.971	0.997	0.982	0.544	0.956	0.870
6	0.742	0.908	0.972	0.997	0.982	0.808	0.961	0.882

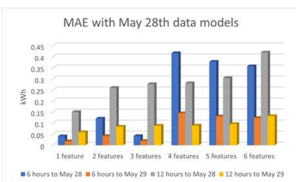
features showed higher error values. In 2(c), the 6-hour models that use 1 or 3 features have lower MAE and MdAE values. Finally, in 2(d) the 6-hour models performed significantly in this case, especially the model that uses a feature for *X*.



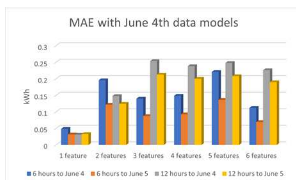
(a) Forecast MAE with hours on May 14^{th}.



(b) Forecast MAE with hours on May 21^{st}.



(c) Forecast MAE with hours on May 28^{th}.



(d) Forecast MAE with hours on June 4^{th}.

Fig. 2**.** Forecast MAE with hours.

For May 14^{th}, 21^{st} and 28^{th}, and June 4^{th}, the R^2 values for the 6-hour and 12-hour models are also shown in Table I. In this case, the 12-hour models of May 14^{th} with 6 features for *X* have higher R^2 values, meaning that *X* explains the variation of *Y* more effectively than in the 6-hour model. Again, it can be noted that in both models of May 21^{st}, the R^2 value exceeds 0.90, showing the ability of the models to fit the samples, with the 12-hour models highlighted. Also, it is noted a low performance for R^2 in the 12-hour models of May 28^{th}, except the 12-hour model with 6 features. The 6-hour models of May 28^{th}, especially those with 5 and 6 features showed significant results for R^2. In addition, emphasis should be given to the 6-hour models of June 4^{th} that showed higher values for R^2, especially the model with 6 features.

After that, adopting the May 14^{th} 12-hour model to predict 31 days after the 14^{th}, the MAE and MdAE values were obtained, as presented in Fig. 3.

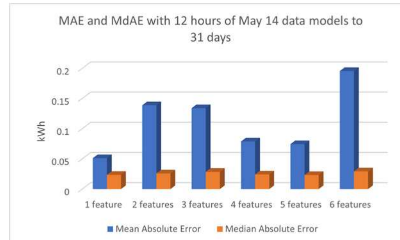
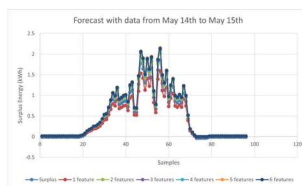


Fig. 3. Forecast MAE and MdAE with 12 hours to predict 31 days after May 14^{th}.

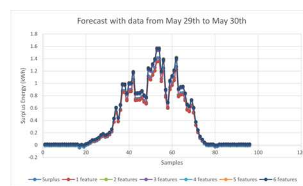
It can be noted that the lowest error values analyzed were obtained from applying the models with 1 feature for *X*.

B. Forecasts from one day to the next

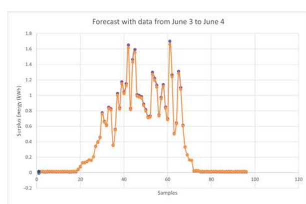
In the context of forecasting a day using the data from the previous day as a basis, this analysis used May 14^{th} to forecast May 15^{th}, May 29^{th} to forecast May 30^{th}, June 3^{rd} to forecast June 4^{th} and May 13^{rd} to forecast June 14^{th}. Fig. 4 presents the predicted data compared to the actual data, called Surplus. The X-axis represents the number of samples.



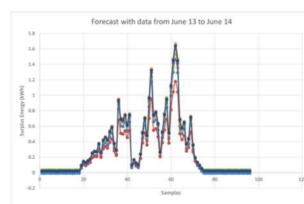
(a) Forecast for May 15^{th} based on May 14^{th} data.



(b) Forecast for May 29^{th} based on May 30^{th} data.



(c) Forecast for June 3^{rd} based on June 4^{th} data.



(d) Forecast for June 13^{rd} based on June 14^{th} data.

Fig. 4. Forecast from one day to the next.

It can be seen that the curves with predicted values are very close to the actual values. Another detail is that in cases where the actual values are negative, the predicted values are null. It is also possible to notice that some predicted value curves present higher errors than others. This can be seen in the following tables.

To improve the analysis of the obtained data, Table II presents the values of ACC, R^2, MAE, MdAE, EVS, and

TABLE II. Models created on May 14^{th} and 29^{th}, and June 3^{rd} and 13^{th} forecasting their next day.

Features	Model created on May 14^{th} Forecast for May 15^{th}						Model created on May 29^{th} Forecast for May 30^{th}						Model created on June 3^{rd} Forecast for June 4^{th}						Model created on June 13^{th} Forecast for June 14^{th}					
	ACC	R^2	MAE	MdAE	EVS	E_{Var}	ACC	R^2	MAE	MdAE	EVS	E_{Var}	ACC	R^2	MAE	MdAE	EVS	E_{Var}	ACC	R^2	MAE	MdAE	EVS	E_{Var}
1	0.69	0.707	0.0686	0.0247	0.976	0.00714	0.88	0.842	0.0211	0.0171	0.997	0.000576	0.91	0.922	0.0139	0.0178	0.999	0.000192	0.74	0.678	0.0574	0.0380	0.951	0.00573
2	0.59	0.707	0.0587	0.0351	0.994	0.00173	0.91	0.851	0.0564	0.0299	0.989	0.00211	0.91	0.923	0.0395	0.0264	0.997	0.000560	0.73	0.699	0.0476	0.0450	0.9997	3.013E-05
3	0.54	0.710	0.0765	0.0386	0.987	0.00392	0.93	0.858	0.0490	0.0305	0.995	0.000982	0.91	0.925	0.0516	0.0299	0.993	0.00153	0.80	0.719	0.0513	0.0404	0.995	0.000534
4	0.54	0.710	0.0794	0.0403	0.986	0.00419	0.89	0.860	0.0569	0.0304	0.990	0.00205	0.92	0.929	0.0694	0.0388	0.985	0.00308	0.78	0.720	0.0633	0.0452	0.988	0.00145
5	0.48	0.711	0.0885	0.0387	0.978	0.00670	0.88	0.862	0.0577	0.0319	0.990	0.00189	0.92	0.931	0.0460	0.0317	0.997	0.000670	0.80	0.726	0.0618	0.0433	0.987	0.00152
6	0.53	0.725	0.0870	0.0413	0.981	0.00571	0.86	0.862	0.0548	0.0305	0.991	0.00171	0.92	0.931	0.0458	0.0316	0.997	0.000658	0.81	0.728	0.0552	0.0394	0.991	0.00111

E_{Var}. For May 14^{th}, the ACC of the model with 2 features for X is lower than with 1 feature, but the MAE is 0.0587 kWh, the lowest among the other models. Another detail is that the ACC of all models does not have high values, which implies that these models may not apply to all new data. Furthermore, it can be seen that the R^2 has low variation among the models and has significant values. For May 30^{th}, the model with 1 feature for X has the lowest MAE and lowest E_{Var} among the other values of these parameters for the other models. In contrast, the highest ACC value corresponds to the model using 3^{rd} features for X. With regard to the R^2, the values for all models are above 0.80, with the model with 6 features for X as the highlight. The lowest MAE value is 0.0139 kWh, which corresponds to the model with 1 feature for X. For June 3^{rd}, the ACC values are above 0.90 for all models, which represents the significant generalization ability of the model to new data. With regard to the R^2 the values for all models are above 0.90, especially for the models with 5 and 6 features for X. Finally, the lowest MAE is 0.0476 kWh, related to the model with 2 features for X, which also has the lowest E_{Var}. In contrast, the ACC of this model is around 0.73 and is the lowest among the other models. It can also be seen that the values for R^2 are lower than the models developed on the other days considered.

In brief, it is possible to note that the ACC values for the models created on the first day are not as significant when compared to the other days. Furthermore, it can be seen that models using 1 or 2 features have lower MAE. The EVS has significant values in both cases and the E_{Var} has relatively low results.

C. Using 5 days for forecast

Continuing the predictive analysis, 5 days of data were selected, between May 14^{th} and May 18^{th}, to predict the sixth day, in this case, May 19^{th}, and the same day 19^{th} in the other months of the year. Fig. 5 presents the actual and predicted data for May 19^{th}, in particular, based on models developed from data from May 14^{th} to 18^{th}

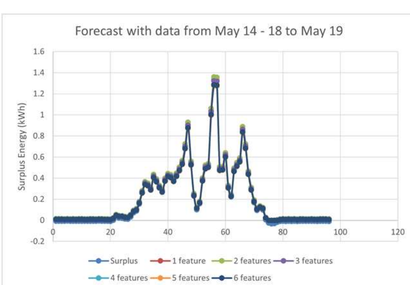


Fig. 5. Forecast for May 19^{th} based on May 14^{th}-18^{th} data.

It is possible to observe a high approximation of the actual and predicted data from May 19^{th}. Fig. 6 presents the MAE and MdAE values for the May 19^{th} forecast and the same May 19^{th} forecast for the following 11 months.

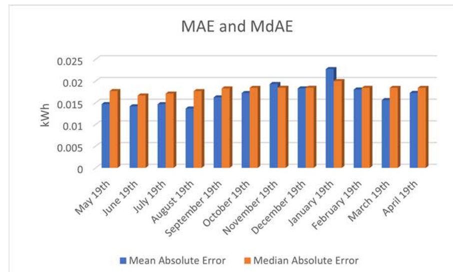


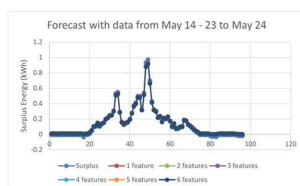
Fig. 6. Forecast MAE with 5 days.

The mean and MdAE values do not differ so much, being higher in the month of January 2017. The E_{Var} is higher in July 2016 and January 2017.

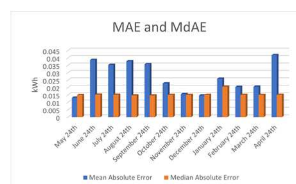
Another parameter to be evaluated is R^2, whose values are higher than 0.845 for all models, especially the model with 6 features whose R^2 is 0.864.

D. Using 10 days for forecast

Using the days between May 14^{th} and 23^{rd} to predict 24^{th}, and the same day 24 in other months, Fig. 6(a) shows the actual and forecast data curves for May 24^{th}. Fig. 6(b) shows the MAE and MdAE values for the May 24^{th} forecast and the same day forecast for the following 11 months.



(a) Forecast for May 24^{th} based on May 14^{th}-23^{rd} data.



(b) Forecast MAE and MdAE with 10 days.

Fig. **7.** Using 10 days for forecast.

It can be seen that the months of May, November, and December are highlighted as having the lowest mean and median errors. In addition, the values found for R^2 are higher than 0.90 for all models, especially the model with 6 features whose R^2 is 0.911, which represents a high proximity of the data to the established regression line.

E. Using 30 days for forecast

Finally, the last analysis use an entire month to predict the day following that month. Using data from May 14^{th} to June

13^{th}, day 14^{th} was predicted. Fig. 8 shows the actual and forecast data curves for June 14^{th}.

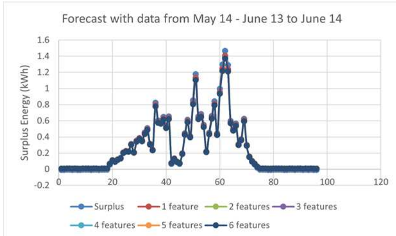


Fig. 8. Forecast for June 14^{th} based on May 14^{th}- June 13^{rd} data.

To complement the results obtained in the graph in Fig. 8, Table III presents the parameters obtained.

TABLE III. Model created on May 14^{th} - June 13^{th} / Forecast for June 14^{th}

Features	ACC	R^2	MAE	MdAE	EVS	E_{Var}
1 feature	0.91	0.907	0.0152	0.0165	0.997	0.000324
2 features	0.91	0.907	0.0202	0.0163	0.994	0.000680
3 features	0.91	0.910	0.0185	0.0134	0.995	0.000584
4 features	0.91	0.910	0.0183	0.0134	0.995	0.000569
5 features	0.91	0.910	0.0184	0.0137	0.995	0.000577
6 features	0.91	0.911	0.0189	0.0147	0.995	0.000604

It can be seen, based on the acquired results, that the values of mean and median error, accuracy, R^2, and EVS showed no significant difference compared to the models with 5 and 10 days. Another detail that can be considered is that, in general, the models that use 6 features for X present higher R^2 values, which means a better performance of these models in explaining the variation of Y.

After the results obtained, it is possible to consider that the objective of this work is to understand how to have a robust forecast model from the amount of data and when to update it. Based on the analysis done in this section, it is possible to conclude that the forecast models with hours or with days proved to be efficient within the experiments performed.

VI. CONCLUSIONS AND FUTURE WORKS

In summary, it is possible to highlight limitations in this work. To understand which data period is best for creating forecasting models, more analysis and experiments can be made, taking into consideration that the longer the time period for model creation, the longer the time the system is not forecasting surplus energy. In addition, as future work, it is intended to perform analysis and prediction with consumption data, and conduct further studies on updating the models, that is, when the models should be updated to continue having the same prediction efficiency and application of system decision-making with the excess power and consumption prediction data.

Furthermore, it is important to highlight that Linear Regression was proposed as a prediction method due to the linear relationship that was observed between the independent variables of the [14] system. However, other ML methods, such as Decision Tree, Random Forest, and several others, can be applied to this system. The suitability of the models depends on the application, which implies that certain methods are widely used for applications that have non-linear relations between their variables. In the case of this work, the linear relationship between the variables made it possible to apply Linear Regression. In this way, as future works and for comparative purposes, it is plausible to mention the possibility of applying other ML methods for the analysis and prediction of data according to the proposal of this work.

ACKNOWLEDGMENT

The authors are grateful to the Foundation for Science and Technology (FCT, Portugal) for financial support through national funds FCT/MCTES (PIDDAC) to CeDRI (UIDB/05757/2020 and UIDP/05757/2020) and SusTEC (LA/P/0007/2021). Thadeu Brito is supported by FCT PhD Grant Reference SFRH/BD/08598/2020 and PhD Student in the Doctoral Program in Electrical and Computer Engineering (PDEEC) of the Faculty of Engineering, University of Porto (FEUP) supervised by Prof. Paulo Costa.

REFERENCES

[1] R. O. Ibrahim, E. Tambo, D. Tsuanyo, and A. Nguedia-Nguedoung, "Modelling an artificial intelligence-based energy management for household in nigeria.," *Engineering Letters*, vol. 30, no. 1, 2022.

[2] M. H. Elkholy, T. Senjyu, M. E. Lotfy, A. Elgarhy, N. S. Ali, and T. S. Gaafar, "Design and implementation of a real-time smart home management system considering energy saving," *Sustainability*, vol. 14, no. 21, p. 13840, 2022.

[3] Z. Ling, J. Luo, Y. Xu, C. Gao, K. Wu, and X. Fu, "Security vulnerabilities of internet of things: A case study of the smart plug system," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1899–1909, 2017.

[4] J. Leitao, P. Gil, B. Ribeiro, and A. Cardoso, "A survey on home energy management," *IEEE Access*, vol. 8, pp. 5699–5722, 2020.

[5] M. R. Sunny, M. A. Kabir, I. T. Naheen, and M. T. Ahad, "Residential energy management: A machine learning perspective," in *2020 IEEE Green Technologies Conference(GreenTech)*, pp. 229–234, 2020.

[6] E. Biondo, T. Brito, A. Nakano, and J. Lima, "A wsn real-time monitoring system approach for measuring indoor air quality using the internet of things," in *Internet of Everything: The First EAI Int. Conf., IoECon 2022, Guimarães, Portugal, September 16-17, 2022, Proceedings*, pp. 76–90, Springer, 2023.

[7] R. G. Rajasekaran, S. Manikandaraj, and R. Kamaleshwar, "Implementation of machine learning algorithm for predicting user behavior and smart energy management," in *2017 Int. Conf. on Data Management, Analytics and Innovation (ICDMAI)*, pp. 24–30, 2017.

[8] S. Asadi, S. S. Amiri, and M. Mottahedi, "On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design," *Energy and Buildings*, vol. 85, pp. 246–255, 2014.

[9] S. Sah, "Machine learning: a review of learning types," *Preprints.org*, 2020.

[10] S. Rong and Z. Bao-Wen, "The research of regression model in machine learning field," in *MATEC Web of Conferences*, vol. 176, p. 01033, EDP Sciences, 2018.

[11] O. P. S. Data, "Data package household data." https://data. open-power-system-data.org/household data/2020-04-15. February, 2023.

[12] M. Holmstrom, D. Liu, and C. Vo, "Machine learning applied to weather forecasting," *Meteorol. Appl*, vol. 10, pp. 1–5, 2016.

[13] A. McQuistan, "Using machine learning to predict the weather: Part 1." https://stackabuse.com/ using-machine-learning-to-predict-the-weather-part-1/. February, 2023.

[14] D. Maulud and A. M. Abdulazeez, "A review on linear regression comprehensive in machine learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, 2020.