Estimation of Power Consumption Prediction of Electricity Using Machine Learning

Paradhasaradhi Yeleswarpu[1], Rakesh Nayak[1](✉), and R. D. Patidar[2]

[1] Department of Computer Science and Engineering, OP Jindal University, Raigarh, India
pard.mt21cs04@opju.ac.in, nayakrakesh8@gmail.com
[2] Department of Electrical Engineering, OP Jindal University, Raigarh, India
rd.patidar@opju.ac.in

Abstract. Electricity consumption has been broadly concentrated on in the PC engineering field since numerous years. While the securing of energy as an action in ML is arising, a large portion of the trial and error is still essentially centered around getting raised degrees of precision with no computational limitations. We accept that one of the reasons for this deficiency of interest is because of their shortfall of straightforwardness with admittance to assess energy utilization. The principal objective of this study is come to assess valuable guidelines to the MLpeople group that grants them the major acknowledgment to utilize and fabricate energy assessment techniques for AI calculations. Utilization of various group models like Linear Regression, and random forest regression and gride search cv, adaboost algorithms to predict the power and to acquire exact outcomes. Notwithstanding, we additionally present the state-of-the-art programming apparatuses that award power assessment standards, along with two use cases that reinforce the request of energy fatigue in ML. Toward the end, we anticipate the future energy which is so useful to the matrix to make exact energy for the network by refreshing with shrewd meters where everyone can know individuals, who are involving more energy in what machines, so it is gigantically useful in which time we want more energy and less energy.

Keywords: Electricity consumption · Random Forest · AdaBoost · Gridsearch CV · Linear Regression · Streamlit tool · Machine Learning

1 Introduction

These days, energy is being utilized further, as a result of the utilization of homegrown and modern purposes, for instance, engine vehicles, enormous scope generators, cell phones, and domestic devices. Moreover, the nonstop development of the framework for savvy meters (SMI) [1]. It was established globally to consolidate dynamic energy frameworks in clever meters. This acquaintance opened the door for gauge or model energy use, and there is currently a chance to apply for a green environment, especially for consumers of domestic energy [2].

P. Pareek et al. (Eds.): IC4S 2023, LNICST 536, pp. 151–162, 2024.
https://doi.org/10.1007/978-3-031-48888-7_13

The use of electrical equipment and consumer behaviour are having an impact on the power industry. The power network organisations are also recognising the necessity to improve and discover better methods to successfully manage power use in contemporary and private structures that control energy interest. While clever private structures provide residents with offices that allow them to operate numerous technological devices substantially through portable applications, the sensors often demand considerable energy usage. A gathering relapse model using the direct prediction and the SVR expectation technique was developed [3] to increase the power expectation's proficiency.

Because of this kind of bad management, household equipment is frequently misused and innumerable assets are lost every year [3]. In order to maintain this energy tragedy by precise interest rates for the foreseeable future, it is especially important to reduce it. A few estimate calculations are used in the energy the board sector to determine power interest in capacity production soon [3]. However, the structure includes a few elements that could affect energy usage, such as the climate, the construction materials, and the sub-level designs for heating, lighting, and ventilation. [4] Customers may alter the pile using machines or tenants, taking into account financial energy use. [5] Basic and dependent on the security and refinement of the structure's framework is the projecting of this energy. Verifiable data with publicly released family values from 2006 to 2010 are used to enable effective use and sending with the expectation of power use. [6].

2 Literature Survey

Corgnati et al. (2013) employed the data (regressor variables) and yield factors (response). This information will be used to evaluate the system boundaries, and as a result, a numerical model might be produced. In a few earlier works, the information driven AI methodology has been explored. Fu et al. (2015) suggested using Backing Vector Machine (SVM), one of ML computations, to predict the load at a structure's framework level (cooling, lighting, power, and others), taking into account weather forecasts and hourly power load input. With a mean predisposition error (MBE) of 7.7% and a root mean square error (RMSE) of 15.2%, the SVM technique accurately predicted the whole power load.

As part of the Brilliant City Demo Aspern (SCDA) project, Valgaev et al. (2016) created a power demand projection utilising the k-Closest Neighbour (k-NN) model at a clever structure. The k-NN gauging method now makes use of a number of verifiable perceptions (daily loadcurves) and their substitutes. Because it only distinguishes between comparable in-positions in a huge component space, the k-NN approach is excellent at organising data but has limitations for predicting future value. As a result, it ought to be strengthened with tenuous information that acts as a sign of expectation for the next 24 h on typical business days.

El Khantach et al. (El Khantach et al., 2019) employed five artificial intelligence (AI) techniques for momentary load anticipating with an underlying disintegration of the real information carried out irregularly into time series of each hour of the day, which finally consisted of 24 timeseries that addressed each preceding hour. The five AI techniques employed are Multi-facet Perceptron (MLP), Support Vector Machine (SVM), Outspread Premise Capability (RBF) Regressor, REPTree, and Gaussian Interaction. Trial and error were carried out in light of the data from the Moroccan electrical burden information. With a MAPE level of 0.96, the results showed that the MLP approach was the most dependable. SVM came in second and, despite performing significantly worse than MLP, was still superior to the other methods.

Expectations could also be created in light of the order-based AI strategy, which is commonly employed for energy usage forecasting, even though Gonzalez-Brioneset al. (2019) concentrated on the relapse technique. The inquiry developed a predictive model by examining the verifiable data collection using Direct Relapse (LR), Backing Vector Regression (SVR), Irregular Timberland (RF), Choice Tree (DT), and k-Nearest Neighbour (k-NN). The exploration's boundaries also contained a further variable known as one-day power usage (kWh). The outcomes showed that, with a score of 85.7%, the LR and SVR models delivered the best correct presentation. The hour cost and apex power-restricting based request reaction approaches serve as the foundation for this planning. They also offered a credible experiment to back up their timetable. The test showed a considerable reduction in the quantity of energy utilized by the various equipment because of the timetable they prepared. Creators in [7] are proposing the development of a home energy the board framework in order to select the optimal day-ahead planning for the various machines.

Sou Family Cheong et al. presented a planning method for intelligent home equipment in light of mixed number straight programming in [8]. They also took into account the machines' usual span and peak power usage. The suggested strategy resulted in cost savings of roughly 47% when compared to a previously specified duty. The authors also showed that with almost minimum computing effort, generally excellent layouts could be created.

There has been a lot of work put into addressing various initiatives to forecast how much energy would be used by different devices in relation to expectations for energy usage. Elkonomou made the expectancy method suggestion in [9] in light of the false brain arrangement. In order to choose the design with the best hypothesis, a number of tests were run using the multi-facet perceptron model. Actual information about the data and outcomes was used during all stages of preparation, approval, and testing.

The importance of the structure's energy usage expectation for the board and efficient energy management is emphasized by the authors of [10]. In order to meet the expectation, they are adopting a model that is information-driven and takes into consideration the forecast for energy use. According to the survey, there are numerous gaps in the field of energy utilization forecasting that need to be solved, including the prediction of long-distance energy use, the prediction of energy used inside of private structures, and the prediction of energy used for structure illumination. This lack of exploration may result from the very scant amount of knowledge that is currently available.

3 Existing System

SVM utilized in the Current arrangement of the issue proclamation. Large informative collections are not a good fit for SVM calculations. When the informational index is more crowded, as is the case when target classes are being covered, SVM doesn't function very well. The SVM won't perform as expected when the number of elements for each information point exceeds the number of information tests that need to be prepared.

4 Proposed Method

Utilizing individual power utilization dataset, We tested the suggested method using a dataset that is freely available from the UCI AI repository and contains details on electricity usage.The Dataset has 198721 lines × 6 sections. We train every irregular timberland calculation and direct relapse and lattice search cv model on the train set utilizing all highlights and afterward assess them on the whole test set. To quantify execution over the long run.

We use the scikit-learn implementation of the following mentods.

1. Random Forest
2. Linear Regression
3. Grid search cv and
4. Adaboost.

The accuracy of neural networks is high if the datasets provide appropriate training. Increasing the accuracy score, Large amount of feature we are taking for the training and testing. The basic architecture is shown in Fig. 1.

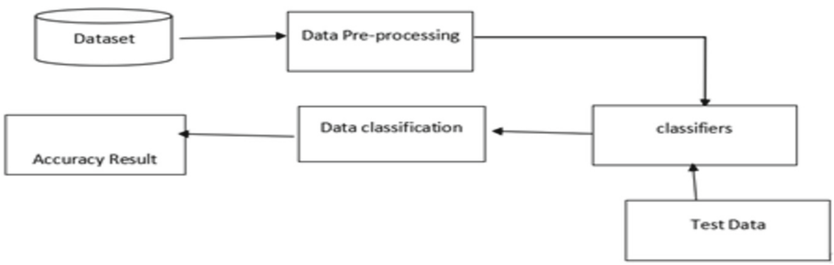

Fig. 1. System Architecture

5 Methodology

Data Gathering,
preprocessing of the data,
feature extraction,
evaluation model, and.
user interface.

5.1 Data Gathering

This paper's information assortment comprises of various records. The determination of the subset of all open information that you will be working with is the focal point of this stage. Preferably, ML challenges start with a lot of information (models or perceptions) for which you definitely know the ideal arrangement. Marked information will be data for which you are as of now mindful of the ideal result.

5.2 Pre-processing of Data

Format, clean, and sample from your chosen data to organize it.

There are three typical steps in data pre-processing:

Designing.

Information cleaning.

Inspecting.

Designing: It's conceivable that the information you've picked isn't in a structure that you can use to work with it. The information might be in an exclusive record configuration and you would like it in a social data set or text document, or the information might be in a social data set and you would like it in a level document.

Information cleaning; is the most common way of eliminating or supplanting missing information. There can be information examples that are inadequate and come up short on data you assume you really want to resolve the issue. These events could should be eliminated. Moreover, a portion of the traits might contain delicate data, and it very well might be important to anonymize or totally eliminate these properties from the information.

Inspecting: You might approach significantly more painstakingly picked information than you want. Calculations might take significantly longer to perform on greater measures of information, and their computational and memory prerequisites may likewise increment. Prior to considering the whole datasets, you can take a more modest delegate test of the picked information that might be fundamentally quicker for investigating and creating thoughts.

5.3 Feature Extraction

The following stage is to A course of quality decrease is include extraction. Highlight extraction really modifies the traits instead of element choice, which positions the ongoing ascribes as indicated by their prescient pertinence. The first ascribes are straightly joined to create the changed traits, or elements. Finally, the Classifier calculation is utilized to prepare our models. We utilize the Python Normal Language Tool stash's classify module.

We utilize the gained marked dataset. The models will be surveyed utilizing the excess marked information we have. Pre-handled information was ordered utilizing a couple of AI strategies. Irregular woodland classifiers were chosen. These calculations are generally utilized in positions including text grouping.

5.4 Assessment Model

Model the method involved with fostering a model incorporates assessment. Finding the model that best portrays our information and predicts how well the model will act in what's to come is useful. In information science, it isn't adequate to assess model execution utilizing the preparation information since this can rapidly prompt excessively hopeful and overfitted models. Wait and Cross-Approval are two procedures utilized in information science to evaluate models.

The two methodologies utilize a test set (concealed by the model) to survey model execution to forestall over fitting. In light of its normal, every classification model's presentation is assessed. The result will take on the structure that was envisioned. Diagram portrayal of information that has been ordered.

5.5 Algorithm

Random Forest

An AI technique called Random Forest is outfit-based and operated. You can combine various computation types to create a more convincing forecast model, or use a similar learning technique at least a few times. The phrase "Irregular Timberland" refers to how the arbitrary woodland method combines a few calculations of the same type or different chosen trees into a forest of trees. The irregular timberland technique can be used for both relapse and characterization tasks. Coming up next are the essential stages expected to execute the irregular woods calculation. Pick N records aimlessly from the datasets. Utilize these N records to make a choice tree. Select the number of trees you that need to remember for your calculation, then, at that point, rehash stages 1 and 2. Each tree in the timberland predicts the classification to which the new record has a place in the order issue. The classification that gets most of the votes is at last given the new record. The Advantages of Irregular Woodland the way that there are numerous trees and they are completely prepared utilizing various subsets of information guarantees that the irregular timberland strategy isn't one-sided. The irregular woods strategy fundamentally relies upon the strength of "the group," which reduces the framework's general predisposition. Since it is extremely challenging for new information to influence every one of the trees, regardless of whether another information point is added to the datasets, the general calculation isn't highly different. In circumstances when there are both downright and mathematical highlights, the irregular woods approach performs well. At the point when information needs esteems or has not been scaled, the irregular woodland method likewise performs well.

Linear Regression

Linear regression Considering how simple the portrayal is, it makes for an appealing model. The response is the anticipated result for the given arrangement of information values (y), and the portrayal is a direct condition that joins that set of information values (x). As a result, both the information value (x) and the result value (e) are numerical. Under the straight condition, each information worth or segment is given one scale variable, known as a coefficient and symbolized by the capital Greek letter Beta (B). The line also receives a second coefficient, commonly known as the catch or inclination coefficient, which increases its level of opportunity (for example, allowing it to completely circle a two-layered map).

For example, the model type in a straightforward relapse situation (one x and one y) would be

$$y = B0 + B1^{*}x \tag{1}$$

When we have more than one piece of information (x), the line is referred to as a plane or a hyper-plane in higher aspects. The condition is depicted by the type of circumstance and the specific characteristics utilized for the coefficients (for example, B0 and B1 in the aforementioned model). The intricacy of a simple relapse model of relapse is frequently discussed. This is a reference to the model's total number of coefficients. When a coefficient hit zero ($0 * x = 0$), the information variable's influence on the model and, subsequently, on the forecast made using the model, is successfully eliminated. This is significant if you consider regularization strategies, which alter the learning calculation to reduce the complexity of relapse models by reducing the overall size of the coefficients and eventually pushing some to zero.

GridSearchCV

In almost every AI project, we train a variety of models on the dataset and choose the one that exhibits the best results. In any event, there is room for improvement because we cannot state categorically that this particular model is the best for the main issue. Our goal is to develop the model in every way possible as a result. One important aspect of these models' presentations is their hyperparameters; by setting appropriate values for these hyperparameters, a model's presentation can be significantly improved. The most popular method for determining the ideal hyperparameter values for a given model is GridSearchCV. As previously said, the value of hyperparameters is crucial to how well a model exhibits. Remember that it is practically impossible to predict in advance which hyperparameters have the finest qualities, thus it is preferable to try all conceivable qualities before deciding which ones are the best. We utilize GridSearchCV to automate the tweaking of hyperparameters because doing it physically might require some effort and resources. The model selection package of Scikit-learn (or SK-learn) has a feature called GridSearchCV. Therefore, it is important to note that we really want the Scikit Learn library to be introduced on the PC. With the help of this capability, you may fit your assessor (model) to your training set and iterate through specified hyperparameters. In the end, selecting the best boundaries from the recorded hyperparameters is possible.

Hyper-boundary tuning alludes to the course of find hyper-boundaries that yield the best outcome. This, obviously, sounds significantly more straightforward than it really is. Finding all that hyper-boundaries can be a subtle craftsmanship, particularly given that it relies generally upon your preparation and testing information. As your information develops, the hyper-boundaries that were once high performing may no longer perform well. Monitoring the outcome of your model is basic to guarantee it develops with the information. One method for tuning your hyper-boundaries is to utilize a Matrix search. This is presumably the least complex strategy as well as the absolute most rough. In a matrix search, you attempt a framework of hyper-boundaries and assess the presentation of every mix of hyper-boundaries. The GridSearchCV class in Sklearn fills a double need in tuning your model. The class permits you to apply a framework search to a variety of hyper-boundaries, and Cross-approve your model utilizing k-overlay cross approval. The interaction pulls a segment from the accessible information to make train-test values. It rehashes this cycle on different occasions to guarantee a decent evaluative split of your information. The capability that takes various boundaries. We should investigate these in somewhat more detail:

Estimator: It takes an assessor object, for example, a classifier or a relapse model.

Param grid: It takes a word reference or a rundown of word references. The word references ought to be key-esteem matches, where the key is the hyper-boundary and the worth are the instances of hyper-boundary values to test.

Cv: It takes a number that decides the cross-approval methodology to apply. On the off chance that None is passed, 5 is utilized.

Scoring: It takes a string or a callable. This addresses the technique to assess the exhibition of the test set.

n_jobs: It addresses the quantity of tasks to run in equal. Since this is a tedious cycle, running more positions in equal (in the event that your PC can deal with it) can accelerate the cycle.

verbose: It decides how much data is shown. Involving a worth of 1 shows the ideal opportunity for each run. 2 shows that the score is additionally shown. 3 demonstrates that the overlap and up-and-comer boundary are additionally shown.

Ada Boosting Classifier

Ada-boost or Adaptive Boosting is one of the help group classifications made by Yoav Freund and Robert Schapire in 1996. It mixes various classifiers to improve classifier precision. AdaBoost is an iterative outfit approach. The AdaBoost classifier builds regions of strength for a, providing you high areas of strength for exactness by combining many classifiers that combine inefficiently. Adaboost's main principle is to set up the classifier loads and get ready for each cycle's information test to the point where it guarantees precise forecasts of unanticipated impressions. The fundamental classifier can be any AI computation that recognizes loads on the training set. Adaboost must abide by two conditions. The classifier needs to be prepared intelligently using a number of weighed preparation models. In order to provide these samples with the greatest fit possible throughout each iteration, it works to decrease training error.

How does the AdaBoost algorithm work? Here is how it works. A training subset is originally selected by Adaboost at random. It iteratively trains the AdaBoost AI model by choosing the preparation set in consideration of the precise expectation of the prior preparation. It gives incorrectly characterized perceptions a heavier burden, increasing their likelihood of grouping in the attention that follows. Additionally, it transfers the burden to the trained classifier in each emphasis in accordance with the classifier's accuracy. The classifier that is more accurate will be given more weight. This cycle repeats until there are the predefined maximum number of assessors or until the entire preparation information fits with virtually minimal error. Play out a "vote" involving all of the artificial learning computations to determine the ranking. The level of precise expectations for the test information is implied by precision. By partitioning the quantity of exact expectations by the complete number of forecasts, it very well might still up in the air.

5.6 User Interface and Result

The pattern of Information Science and Examination is expanding step by step. From the information science pipeline, one of the main advances is model sending. We have a ton of choices in python for sending our model. A few well-known systems are Carafe and Django. Yet, the issue with utilizing these systems is that we ought to have some information on HTML, CSS, and JavaScript. Remembering these requirements, Adrien Trouville, Thiago Teixeira, and Amanda Kelly made "Streamlit". Presently utilizing streamlit you can send any AI model and any python project easily and without stressing over the frontend. Streamlit is very easy to use.

In this article, we will get familiar with a few significant elements of streamlit, make a python project, and convey the task on a nearby web server. How about we introduce streamlit. Type the accompanying order in the order brief.

pip installs streamlit.

When Streamlit is introduced effectively, run the given python code and in the event that you don't get a mistake, then streamlit is effectively introduced and you can now work with streamlit. Figure 2. Shows the user interface to execute the code.

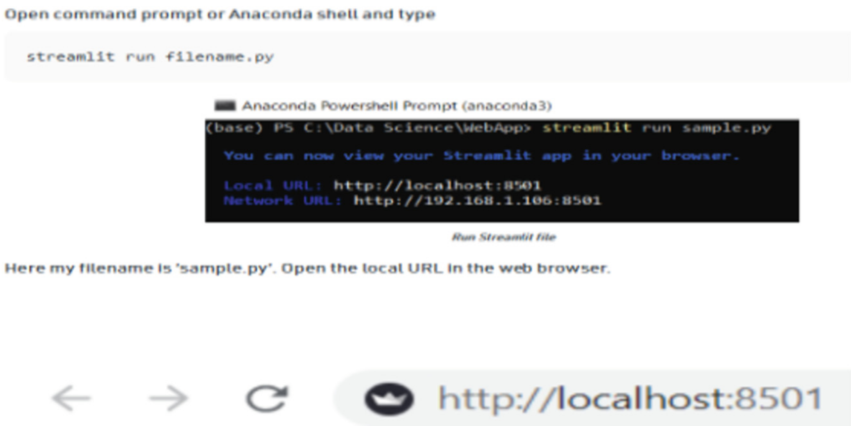

Fig. 2. User Interface

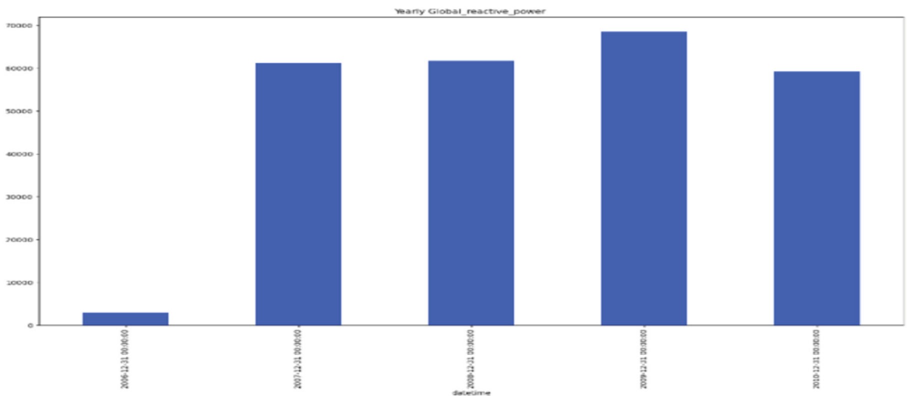

Fig. 3. Bar Chart of yearly Global Reactive Power

Figure 3. Shows bar chart for yearly global reactive power consumption.

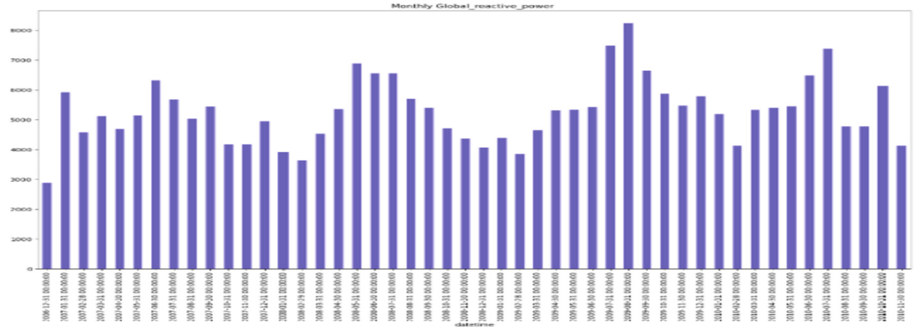

Fig. 4. Monthly Global reactive Power Bar chart

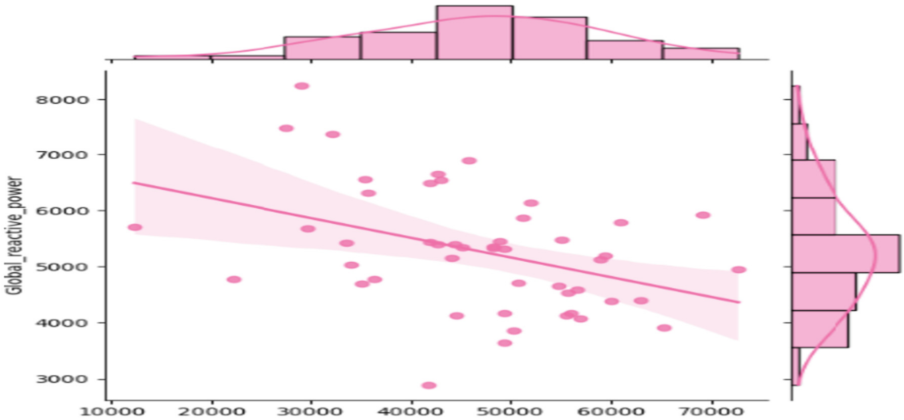

Fig. 5. Global Reactive Power and Distribution Chart

Figure 4. Shows a bar chart for monthly global reactive power consumption.

Figure 5. Shows scatter chart for monthly global reactive power consumption and distribution.

Figure 6. Shows monthly global reactive power consumption.

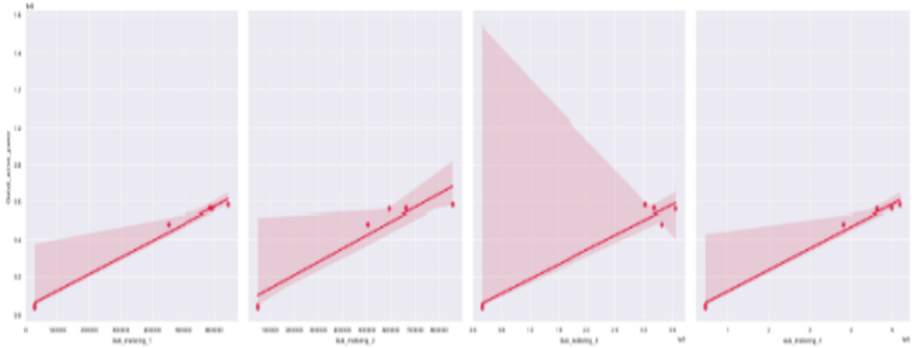

Fig. 6. Subplots of Global Active Power

6 Conclusion

Machine learning (ML) techniques has as of late contributed very well in the headway of the expectation models utilized for power utilization. Such models profoundly work on the exactness, heartiness, and accuracy and the speculation capacity of the ordinary time series anticipating instruments. By using the historical data we can predict future power consumption. Here, we used the electric power consumption data of one household and applied linear regression, ada boost, grid search cv and random forest algorithms and we achieved an linear regression accuracy with 99% and applied random forest algorithm with 93% accuracy, and applied adaboost algorithm with 97% accuracy and we achieved an grid search cv accuracy with 82% for future prediction.

References

1. Corgnati, S., Bednar, T., Jang, Y., Yoshino, H., Ghiaus, C.: Total Energy Use in Buildings. Analysis and Evaluation Methods, Final Report Annex 53. Statistical Analysis and Prediction Methods (2013)
2. Fu, Y., Li, Z., Zhang, H., Xu, P.: Using support vector machine to predict next day electricity load of public buildings with sub-metering device. Procedia Eng. **12**, 1016–1022 (2015)
3. Valgaev, O., Kupzog, F., Schmeck, H.: Low-voltage power demand forecasting using K-nearest neighbors approach. In: IEEE Innovative Smart Grid Technologies - Asia (ISGT - Asia), pp. 1019–1024 (2016)
4. El Khantach, A., Hamlich, M., Belbounaguia, N.E.: Short-term load forecasting using machine learning and periodicity decomposition. AIMS Energy **7**(3), 382–394 (2019)
5. González-Briones, A., Hernández, G., Corchado, J.M., Omatu, S., Mohamad, M.S.: Machine learning models for electricity consumption forecasting: a review. In: IEEE Virtual-Ledgers-Tecnologías DLT/Blockchain y Cripto-IOT (2019)
6. Paterakis, N.G., Erdinç, O., Bakirtzis, A.G.: Catalão JPS Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies. IEEE Trans Ind Inform **11**(16), 1509–1519 (2015)
7. Cheong, S.K.: Scheduling smart home appliances using mixed integer linear programming. In: 50th IEEE conference on decision and control and European control conference, IEEE. https://doi.org/10.1109/cdc.2011.6161081(2011)

8. Ekonomou, L.: Green long-term energy consumption prediction using artificial neural network. Energy **35**(12), 512–517 (2010)
9. Amasyali, K.: A review of data-driven building energy consumption prediction studies. Renew Sustain. Energy Rev. **81**(11), 1192–1205 (2018)
10. Li, Q., Peng, C., Chen, M., Chen, F., Kang, W., Guerrero, J.M.: Networked and distributed control method with optimal power dispatch for islanded microgrids. IEEE Trans. Ind. Electron. **64**, 493–504 (2016)
11. Kong, W., Dong, Z.Y., Hill, D.J., Luo, F., Xu, Y.: Short-term residential load forecasting based on resident behaviour learning. IEEE Trans. Power Syst. **33**, 1087–1088 (2017)
12. Khan, P.W., Byun, Y.C., Lee, S.J.: Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources. Energies **13**, 4870 (2020)
13. Ahmad, A., Hassan, M., Abdullah, M., Rahman, H., Hussein, F., Abdullah, H.: A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew. Sustain. Energy Rev. **33**, 102–109 (2014)
14. Zhang, M., Bai, C.: Exploring the influencing factors and decoupling state of residential energy consumption in Shandong. J. Cleaner Prod. **194**, 253–262 (2018)