Educational Administration: Theory and Practice
2024, 30(5), 5525-5534
ISSN: 2148-2403
https://kuey.net/
Research Article

Appliances Energy Prediction Using Random Forest Classifier

Shan e Fatima[1], Sandeep Kumar[2], Tasleem Jamal[3]

[1,3] Assistant Professor, Dept.of Computer Science and Engineering
[2] Student, M. Tech CSE AI & ML
[1,2,3] Khwaja Moinuddin Chishti Language University, Lucknow, U.P, India shanefatima@kmclu.ac.in shan.ftm@gmail.com

Citation: Shan e Fatima (2024), Appliances Energy Prediction Using Random Forest Classifier , *Educational Administration: Theory and Practice*, *30*(5), 5525-5534
Doi: 10.53555/kuey.v30i5.3814

ARTICLE INFO

ABSTRACT

The objective of this work is to simplify the decision-making process that energy providers must go through in order to decide whether or not to provide a range of residential buildings with energy depending on the demand for that energy. Using the Multi-layer Perceptron model and the Random Forest model, respectively, we classify residential structures in this article according to the amount of energy that they use.

In order to keep track of the temperature and humidity levels within the house, a network of wireless sensors powered by ZigBee was used. There was an interval of about 3.3 minutes between the transmission of the temperature and humidity measurements by each wireless node. After that, an average was calculated using the wireless data based on intervals of ten minutes. Every ten minutes, the data on the energy consumption was recorded using m-bus energy meters. Using the date and time column, the weather information from the weather station at the closest airport, which was Chèvres Airport in Belgium, was obtained from Reliable Prognosis (rp5.ru).

The retrieval of the data, the extraction of the features, and the prediction are the three processes that make up the prediction. During the data retrieval process, the database is queried to obtain the data that has been utilized hourly and on a daily basis. Statistical characteristics such as "mean," "standard deviation," "skewness," and "kurtosis" are derived from the data that was obtained through the use of feature extraction. At this point in the process, the prediction stage, Multi-layer Perceptron and Random Forest have been used to anticipate which buildings would have high or low electricity usage.

Keywords: Wireless Sensors, Zigbee, Multi-layer Perceptron, Random Forest Classifier.

Introduction

The management of electricity in today's sophisticated power networks is fraught with a great deal of difficulty. The "smart grid" is an innovative idea that was developed in order to fully automate the apparatus that is in charge of regulating the regulation of the power supply. This was done in order to make the system more efficient. This became done with the purpose of finding solutions to the problems that have been stated. The intelligent grid is the integration of all of the components that comprise the power system, such as sensors, actuators, controllers, and computing capabilities. This may be characterized as the integration of all of the elements that make up the power machine. This integration is carried out with the purpose of enhancing the general performance of the characteristics that are intrinsic to the energy structures that are already in place. The smart grid is an application of current technology that enables significant changes to be made in the operations, maintenance, and planning of the whole energy management system. These advancements may be made possible thanks to the smart grid. Utilizing cutting-edge generation allows for the addition of such advancements wherever they are needed. This, in turn, leads to increased control over the amount of energy that is utilized, which, in turn, results in lower pricing. The preservation of contemporary communication networks was the responsibility of a number of different governments. This was done in order to reduce the amount of power that is used and to maintain the integrity of the environment's components.

The United States Department of Energy has provided many explanations of the obligations that need to be met by a smart grid in order for it to function properly. These tasks are listed in order of importance.
The smart grid needs to be able to perform a number of important functions, the most important of which are the ability to heal itself, the ability to deter attacks, the provision of power of a high quality, and the ability to enable energy markets to effectively manage energy in accordance with consumer demands. The smart grid is a key component in the effort to guarantee that energy is produced in an efficient manner and in accordance with the needs of end users. The energy market acts as a conduit that connects the many parties involved in the production, distribution, and use of electrical power: consumers, suppliers, and generators. In order for energy providers to get favorable rates for stored energy, short-term energy forecasting, which may include daily and even hourly forecasting, must be performed with a high degree of accuracy. The residential sector accounts for a disproportionately large share of total energy use, which is why most energy providers concentrate their efforts on this market. The residential sector in European nations used a significant quantity of electricity in 2007, and ever since that year, the residential sector has received more attention in comparison to the preceding figures. In 2007, the residential sector consumed a considerable amount of power in European countries. In order to secure favorable rates for traded energy, the power providers involved need to make accurate projections on the amount of energy that will be used the next day.
In this article, we categorize residential buildings according to the amount of energy that they consume by using the Multi-layer Perceptron and Random Forest models respectively. It has been possible to make educated guesses about the hourly power consumption histories of two distinct types of buildings: those with high power consumption and those with low power consumption. The retrieval of the data, the extraction of the features, and the prediction are the three processes that make up the prediction.

Primary Goals of the Thesis

1. Conduct an investigation into the domestic energy usage records of the past:
One is able to highlight the behavior of the individual collections in addition to the interdependence between wonderful collections by performing a statistical analysis on a given historic time series of power intake from individual homes. This allows one to highlight both the behaviours of the individual collections as well as the interdependence between wonderful collections. This will be accomplished by carrying out an analysis on a certain historic time series that has been provided.
2. Evaluate available mathematical models and choose those that are applicable:
Examine and evaluate mathematical models, with special emphasis on their capacity to accurately represent load dynamics and predict the load that will occur in the future.
3. Experiment with different sizes of clusters and different types of grouping:
Carry out several tests to determine the effect that clustering has on the accuracy of the load prediction, as well as the size of the clusters.
4. Discuss the consequences for the load projections for the immediate future:
Discuss the advantages and disadvantages of the models that have been presented, taking into account the findings from the points made above, and suggest modifications to make it possible for future studies to improve research within STLF.

Scope and Limitations

The Pecan Street Data Set, which is part of a research project undertaken in the Austin region in Texas with the purpose of addressing the global water and energy crisis, is used in this thesis. The study project's primary location is Texas. As a result of this, any conclusions that are drawn may only be applicable to the population that was researched in Austin, and extrapolations to other geographic locations need to be done with caution. Having said that, a large number of research are now being conducted on STLF using a variety of model settings applied to a wide range of datasets. As practitioners test established model configurations on fresh datasets in a variety of geographical locations, the capacity to generalize develops as a result of this testing.
Support Vector Regression and Random Forest are the models that are used for the purpose of forecasting the future load, where the fundamental cause for picking each model is defined. A baseline model that also functions as a benchmark is developed so that the quality of the models can be evaluated. Before constructing a model for forecasting, the K-Means Clustering algorithm is used to group homes according to the average daily load profile of each family. This is done with the intention of identifying the influence that may be had by grouping households that are similar. All of the computations were carried out using Python 2.7 and jupyter, utilizing the foundational algorithm implementations provided by scikit-learn.

Related Work on the Proposal

There are a lot of different techniques to predicting energy consumption that can be found in the research literature, but the bottom-up approach is the most crucial one. When using a bottom-up technique, the first step is to make a prediction of the amount of energy utilized by each particular household appliance. The overall amount of energy that will be used in the smart home is estimated after adding up the amount of energy that was expected to be used by each individual device. The estimation of the amount of electricity used by particular home appliances on an hourly, daily, weekly, and monthly basis is also quite substantial. However,

the prediction of the combined total power intake in the intelligent home is extremely important for the environmentally friendly control of power supply to the residential area.
The purpose of this study is to classify residential homes into one of a kind businesses, high power consumption residential buildings, and coffee strength residential homes in line with the quantity of electricity that is fed on by means of the people who live in every of these exceptional kinds of buildings. This categorization of residential homes into one of a kind businesses, high power consumption residential buildings, and coffee strength residential homes is based at the findings of this study. It is possible to draw parallels between the two of these organisations. The forecast may be of assistance to power providers in making informed decisions regarding the fulfilment of the power requirements in the homes and businesses of their customers. In a similar vein, having access to this projection will benefit the control of structures for smart home automation.
The authors have proposed a structure for the intelligent home power management device, which consists of three layers of different materials. These three layers are referred to by their individual names, which are the anticipative layer, the neighborhood layer, and the reactive layer respectively. The anticipative layer is the location where competitions for sports that entail prediction, such as price prediction, climate prediction, and energy prediction, are held. Some examples of these types of sports include making predictions about the weather, the energy market, and the prices of other commodities. The data that relates to the prediction is sent to the reactive layer in order to make the functioning of that layer more effective. This is done in order to improve the accuracy of the forecast.
As the information travels over the miles, it is first presented to the local layer, which is immediately introduced from the reactive layer. according to the information gained from the reactive layer, the only layer that is responsible for ensuring that the functions of each and every piece of domestic home equipment are managed in a way that is suitable for the requirements of the user is the nearby layer. The fourth layer of the model is the one that is responsible for managing any energy that enters the system from outside the system. This requirement is included within the scope of its authority. The power providers get their supply of energy from the external market in this stratum. The energy used to supply the external market originates from a variety of sources, including nuclear power plants, hydropower plants, thermal power plants, and renewable resources.
In the body of research that has been done, several kinds of classifications and predictions have been made using a variety of different classifiers. Both categorization and prediction were accomplished by the authors via the use of Bayesian networks.

Table 1. Accuracy table of Authors

Noteworthy Contributions by various Authors with Accuracy Rate			
S.no	**Author**	**Published Year**	**Accuracy Rate**
1	Ardwin Kester S.Ong	2022	93.44%
2	Reny Nadifatin	2022	97.00%
3	Satria Fadil Persada	2022	94.85%
4	Muhammad Aslam	2022	66.45%
5	Manar Amayri	2022	95.86%
6	Dorin Maddovan	2021	93.44%
7	A Slowik	2021	97.00%
8	Thais Berrettini Basco	2021	98.56%
9	TKS Shavali	2021	97.89%
10	Ziwei Xiao	2021	88.78%
11	Xin Wv	2020	83.44%
12	Dian Jiao	2020	93.44%
13	Ejaz Ul Haq	2020	98.2%
14	Michel Chammas	2019	72.82%
15	Abdallah Makhow	2019	61.48%

Research Methodology

When an honest kind of prediction is utilised, which is to say when historical records are consulted, the classification method is used in order to make forecasts. In this work, we have also carried out the procedure of forecasting the amount of power that was used based on the model technique. As a finished component of the total paintings, this was completed. The strategy that is now under discussion categorises residential structures as either high-energy intake residential buildings or low-energy intake residential buildings, depending on the total amount of electricity that the buildings use. With the aid of the suggested strategy, the power wholesaler could be able to improve their ability to prepare for the potential power requirements of residential structures. The classification was finished with the assistance of the Multi-layer Perceptron and the Random forest, and a total of four characteristics were thought about throughout the process. All of the functions' most important statistics, including the mean, the standard deviation, the kurtosis, and the skewness, are hidden away inside of them. In order to carry out the accuracy evaluation, the information from the Confusion Matrix and the Kappa statistic have been used.

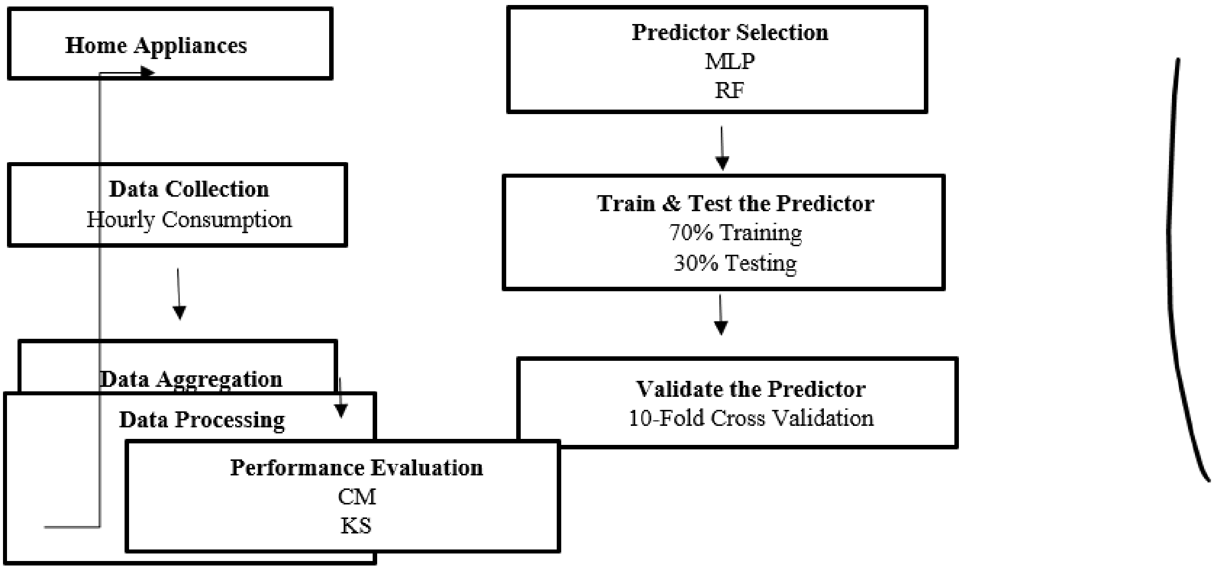

Figure 1. Methodology of the Process

Dataset Description

The time interval for the data set has been set to 10 minutes over the last 4.5 months. In order to keep track of the temperature and humidity levels within the house, a network of wireless sensors powered by ZigBee was used. There was an interval of about 3.3 minutes between the transmission of the temperature and humidity measurements by each wireless node. After that, an average was calculated using the wireless data based on intervals of ten minutes. Every ten minutes, the data on the energy consumption was recorded using m-bus energy meters. Using the date and time column, the weather information from the weather station at the closest airport, which was Chievres Airport in Belgium, was obtained from Reliable Prognosis (rp5.ru). This information was then integrated with the experimental data sets. Two random variables are included in the data set so that regression models may be tested and non-predictive characteristics can be eliminated (parameters).

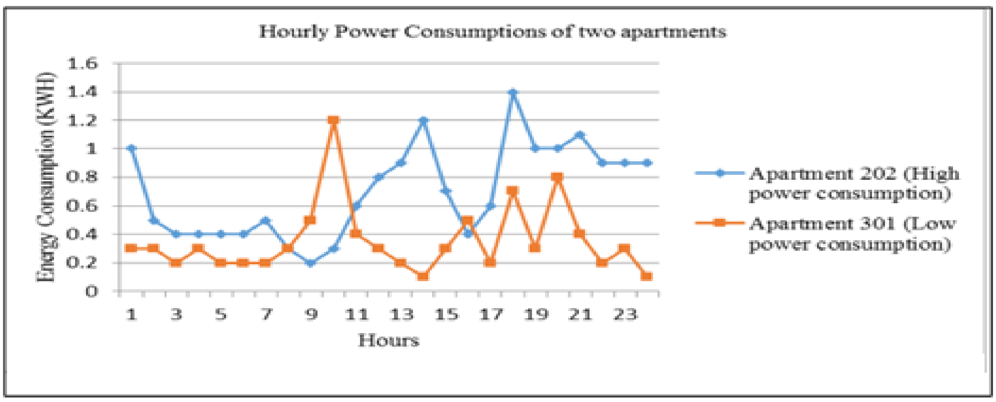

Figure 2. Hourly Power Consumption of two Apartments

Data Processing

The information, together with the intake split up into hours, will be processed using the method that has been selected. It has been established what the mean, standard deviation, and skewness of the data on hourly intake are. This information was gathered. The following is an exhaustive overview of those characteristics:

Mean

The term "mean" refers to the average amount of energy used over the course of twenty-four hours, which may be determined by (1). The letter M stands for the mean of all of the hourly amounts of electricity that were utilised during the whole day. The value of I may range from 0 to 23, and it represents the amount of electricity used during the ith hour of the day. The letter N stands for the total number of hours, which is 24.

$$\boldsymbol{M} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \textbf{(1)}$$

Variance

The term "variance" refers to the differences that occur in the hourly power consumption during the whole day, which may be symbolised by the letter "V." and calculated by (2).

$$\boldsymbol{V} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \qquad \textbf{(2)}$$

Skewness

Skewness is a measure of the imbalance in the hourly power consumption over the course of the full day, and it is denoted by the letter S and calculated by (3).

$$\boldsymbol{S} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3 \qquad \textbf{(3)}$$

Kurtosis

The frequency of very high hourly power consumption by domestic appliances over the course of a whole day is denoted by the letter K and is referred to as kurtosis and calculated by (4).

$$\boldsymbol{K} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4 \qquad \textbf{(4)}$$

Procedure

- Load the data
- Further perform techniques like plotting the energy consumption with respect to time
- Applying features selection techniques on the following dataset and selecting the features like ['low_consum','high_consum','hours','t6','rh_6','lights','hour*lights', 'tdewpoint',' visibility',' press_ mm_ hg',' windspeed']
- We split the data into train and test of 80-20 %
- Later predicting the energy consumption using the above feature set using the Random Forest model with the help of Randomized Search cv where the n_estimators in random forest varies from 150-200

A Classifier for the Purpose of Prediction

The information that was contained in the Excel sheet that tracked the hourly quantity of energy that was utilized was obtained. After the statistics had been obtained, the first step in the processing section was to compute the mean, standard deviation, and skewness of the data. This was followed by the next stage, which was to determine the skewness of the data. The analysis of the data has been completed, and it is now possible to generate the prediction by using those records. The predictor will divide the flats into two separate groups: those that have a low overall electricity usage and those that have a very high overall electricity consumption. For the purpose of this inquiry, Multi-layer Perceptron and Random woodland were applied as approaches to do the work of forecasting the amount of energy that turned becoming used.

Multi-layer Perceptron

In this investigation, we made use of a Multi-layer Perceptron, often known as an MLP, model that had input, hidden, and output layers. The algorithm that is often referred to as the perceptron is one that is supervised and converts one set of input to another. Because it is a linear classifier, the classification is determined by the linear predictor function. This implies that the linear predictor function is in charge of the classification. In order to construct a classification, this function combines the vector that was provided as input with a number of weights. In order for it to be able to solve computational issues, the Perceptron method makes use of a number of layers, each of which is connected to the other levels by a directed network. Each subsequent layer is tightly entwined with the one that comes immediately before it. In order to achieve the results that were aimed for during the whole of the MLP training process, the back-propagation approach was used.
The Perceptron is capable of deriving a single output from a number of different inputs by performing a linear combination in accordance with the input weights and a nonlinear activation function that is calculated by (5). Where w stands for the weighted vector, x stands for the input vector, Y stands for the output vector, b stands for biasedness, and represents the activation function.

$$\boldsymbol{Y} = F(x) = \varphi\left(\sum_{j=1}^{n} \omega_j x_j + b\right) = \varphi(w^T x + b) \quad \textbf{(5)}$$

The architecture of the three layer Multi-layer perceptron is:-

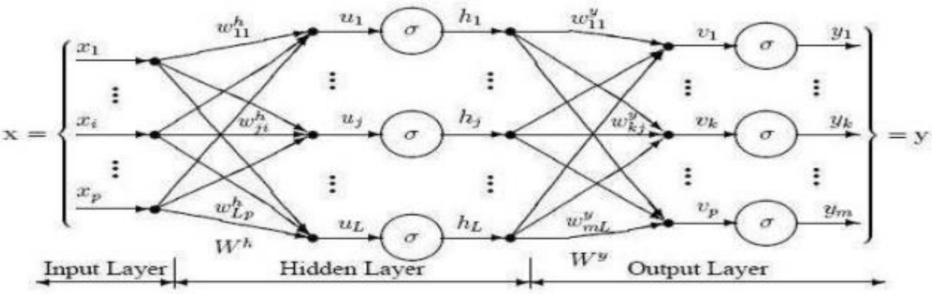

Figure 3. Architecture of Multi-layer Perceptron

Random Forest

Random forests are a form of ensemble learning technique that may be used for a variety of various tasks, including class and regression analysis. Random forests, which are also known as random choice forests, derive their name from the fact that the training phase of the method involves the creation of a large number of different decision bushes. This is how random forests got their name. A subclass known as random desire forests is included in the class that is simply referred to as random forests. The outcome of the random forest is the categorization of the forest that is chosen by the majority of the bushes that are found inside the forest. This is the output of the random forest when it comes to issues regarding type.

There is a technique for supervised learning that is often spelt with an apostrophe, and it is called Random Forest Regression (sometimes spelled Random Forest). When doing regression analyses, the ensemble learning approach is used for the purpose of producing the best possible results. The ensemble learning method is a method that combines the predictions from various machine learning algorithms in order to generate a prediction that is more accurate than a single version with the assistance of Randomized search CV as randomized search cv enables us to conduct some move validation. The goal of this method is to generate a prediction that is more accurate than a single version. The ensemble learning technique is a method that combines the predictions that can be generated with the aid of numerous different device learning algorithms to get a forecast that is more accurate than a prediction generated through a single model. This is accomplished in order to obtain a forecast that may be used to make decisions.

An ensemble classifier known as the Random forest is used throughout the process of determining styles and allocating categories to the many patterns that are identified. That brings the total number of classification schemes that we've used up to two. It is feasible to consider it both as a classifier that is made up of several exceptional classification strategies or as a classification technique that contains a variety of parameters. Both of those perspectives are viable. Each viewpoint has merit in its own right. Supposing that we want to study something, and that we have a mastering set L = ((m1, n1),..,(mi, ni)) with I vectors containing mX and the big apple, in which X represents the observations and Y represents the splendor labels, shall we say that we have a studying set L = ((m1, n1),..,(mi, ni)), and permit's additionally assume that we have The new instance is designated in accordance with each and every tree in the forest on an individual basis according to their character.

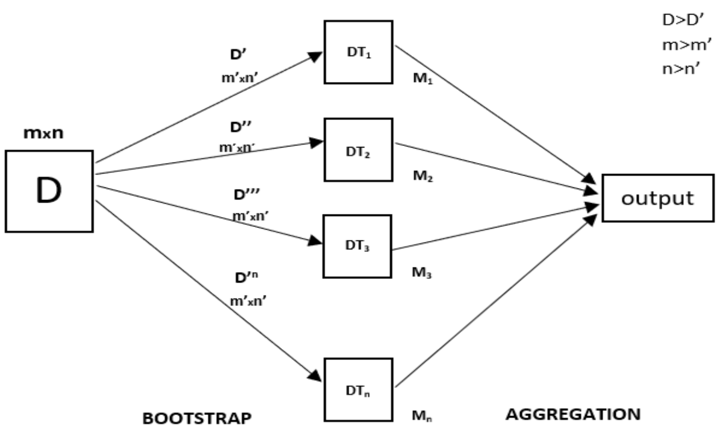

Figure 4. Structure of the Layer

Confusion Matrix

The confusion matrix is a table that compares the outcomes that have been predicted by the prediction device to the results that have been clearly attained. It does this by highlighting any discrepancies between the two sets of information. The term "confusion matrix table" may also be used to refer to this table. When assessing the degree to which the prediction device is correct, the entries included within the matrix are taken into consideration as part of the procedure. The times of the anticipated training may be both "high-quality" and "poor," which would imply that the rental that is the use of the electricity may additionally either belong to the excessive electricity intake apartment or the low electricity intake condo depending on the quantity of electricity it's far consuming. The times of the predicted training may additionally both be "high-quality" and "poor." An explanation of each entry that constitutes the confusion matrix that this take a look at provides is provided in the following paragraphs:

1. The value of A reflects the total number of accurate predictions made for negative events.
2. The value of B reflects the total number of erroneous predictions made for positive situations.
3. The value of C indicates the total number of incorrect predictions made for negative situations.
4. The value of D indicates the number of successful predictions made for positive situations.

Table 2. Predictions of the Data

	Predicted	
	Negative	**Positive**
Actual	A	B
	C	D

The percentage of the total number of right predictions that the same may be and is computed using to determine how accurate a prediction is is known as the accuracy of the forecast (6).

$$\textbf{Accuracy} = \frac{A+D}{A+B+C+D} \qquad (6)$$

Result and Discussion

The model that performed the best, a Random Forest with K = 32 clusters, suggests that the most effective way to group the households based on the daily load profile of each household is to divide the 187 households into 32 distinct clusters. This is the recommendation that comes from the model that performed the best. The clustering of homes inside the 32 clusters is shown in Appendix C. Based on this information, we may deduce that there are essentially ten major groupings of typical households, followed by a number of houses that are highly specialized. Grouping households according to the average daily load profile of each household is undeniably an interesting and intuitive way to cluster households. However, it would be interesting to see how using other representations to cluster households affects the outcome of the predictions to see how using other representations to cluster households would be interesting.

Figure 4.1. Plotting of the Appliances part-1

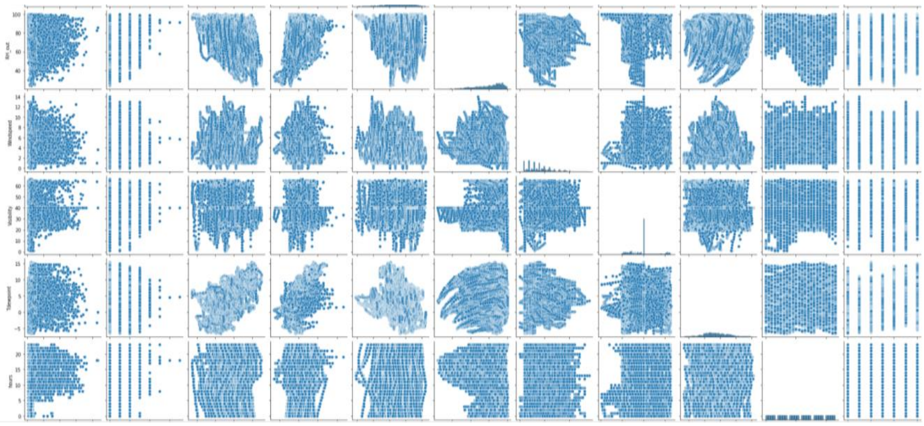

Figure 4.2 Plotting of the Appliances part-2

A broad pattern can be deduced from the findings as follows: Random Forest demonstrates a performance that is consistent across all cluster sizes and exceeds the baseline in every instance for which it was evaluated. SVR, on the other hand, although in certain instances it performs better than Random Forest, it begins to diverge as the cluster size grows and is ultimately outperformed by the baseline. The tests were carried out several times using a variety of seeds, however Random Forest did not exhibit any appreciable differences in its overall performance each time. In addition, since the effectiveness of SVR is unaffected by the element of chance that is included into the forecasting process, any fluctuations in the data are eliminated from the findings. Random Forest produces the most accurate findings when applied to the problem of predicting individual homes, followed by the baseline model in terms of desirability.

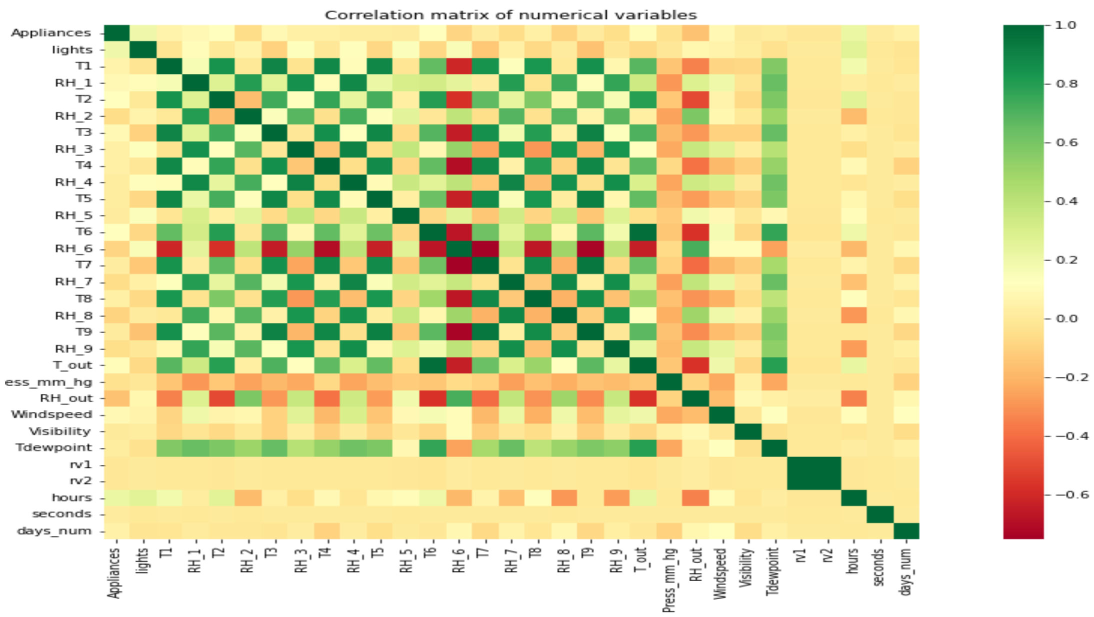

Figure 4.3. Correlation matrix of numerical variable

Table 2: Evaluation of the numerical variables

```
] evaluate(model_cbr, X_test, y_test)

<catboost.core.CatBoostRegressor object at 0x000001D5972022B0>

Average Error        : 11.2618 degrees
Variance score R^2   : 65.77%
Accuracy             : 80.97%
```

When using this methodology, the number of clusters that were applied to the aggregate model in order to provide it with more information had a substantial impact on the results of the predictions. This new technique relates grouping to a lower number of clusters, which is in direct opposition to the ideal cluster-size that should

be used for cluster-based forecasting. The Random Forest model performed better when compared to the aggregate model, and the best model was achieved when information was added based on the grouping into K = 3 clusters, which led to predictions based on 20 characteristics. Random Forest also performed better than the aggregate model. This result is likewise equivalent to the results of the cluster-based forecasting, despite the fact that it does not represent an improvement. When additional information was given to the model, SVR did not demonstrate any meaningful improvement; rather, the results rapidly deteriorated in comparison to the baseline model. One of the underlying causes for the decrease in SVR might be comparable to something that was only recently mentioned, and that is the possibility that the model could need an extensive parameter search. It's possible that when more information is provided, the number of features climbs to the point where considerable overfitting takes happen. This would be another explanation why the performance of Random Forest and SVR stays the same as more information is added. When information from K = 32 clusters is added to models, the resulting models include 136 features, as opposed to merely eight features when the information was added initially.
We achieve a test accuracy of 0.978 and a train accuracy of 0.99

Conclusion and future work

The primary objective of this study is to develop a model for the prediction of the amount of energy that is used by residential flats. This model will assist the system in better organizing the amount of energy that is produced and consumed. Additionally, it will be helpful in determining if residential structures will have a high or low rate of energy consumption (energy management). This could also be helpful for the system that manages the electricity in calculating how much of a daily charge to apply to the energy payments of different units depending on the quantity of power that is consumed by each apartment. The process of making a prediction can be broken down into three levels: retrieving the statistics, determining the characteristics, and actually making the prediction itself. Both the Multi-Layer Perceptron and the Random woodland models have been utilized as predictive tools in the process of selecting the residences that will have a high energy intake and the residences that will have a low energy intake.
The MLP has been shown to be more successful than the Random forest in this one-of-a-kind study that has been done. These studies have been carried out with the intention of determining whether or not it's miles viable every day estimate the quantity of electricity that is used by means of domestic home equipment based totally at the relative humidity and temperature of the home.
The work that desires to be completed inside the future will consist of including greater predictive parameters, which include records approximately occupancy, the vicinity of the house, activities carried out with the aid of the occupant, additional indoor and outdoor surroundings conditions, and the recognition that the usage of an expansion of datasets also can make the prediction greater accurate. This work is anticipated to be completed inside the future. This could be finished in an efficient manner. Even though the use of power is a major problem, it still has a great deal of unrealized potential for the future. The training of extra machine fashions, such as ANN (artificial Neural network), may also further increase the power daily count on.

References

1. C. W. Gellings, "The smart grid: enabling energy efficiency and demand response", The Fairmont Press, Inc., (2009).
2. A. Ipakchi and F. Albuyeh, "Grid of the future," Power and Energy Magazine, IEEE, vol. 7, (2009), pp. 52-62.
3. M. C. Williams, J. P. Strakey, and W. A. Surdoval, "The US department of energy, office of fossil energy stationary fuel cell program," Journal of Power Sources, vol. 143, (2005), pp. 191-196.
4. D. Tren, "EU energy and transport in figuresâ€“Statistical pocketbook 2009, Directorate-General for Energy and Transport," Office for Official Publications of the European Communities, Luxemburg, (2009).
5. D. L. Ha, S. A. Ploix, E. Zamai, and M. Jacomino, "Realtimes dynamic optimization for demand-side load management," International Journal of Management Science and Engineering Management, vol. 3, (2008), pp. 243-252.
6. K. Basu, L. Hawarah, N. Arghira, H. Joumaa, and S. Ploix, "A prediction system for home appliance usage," Energy and Buildings, vol. 67, (2013), pp. 668-679.
7. N. Arghira, L. Hawarah, S. P. Ploix, and M. Jacomino, "Prediction of appliances energy use in smart homes," Energy, vol. 48, no.1, (2012), pp. 128-134.
8. L. Hawarah, M. Jacomino, and S. P. Ploix, "Smart Home-From User's Behavior to Prediction of Energy Consumption," ICINCO 2010, Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics, Volume 1, Funchal, Madeira, Portugal, June 15-18, (2010).
9. I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann, (2005).
10. Wahid, F., R.Ghazali, M.Fayaz and A.S.Shah, "A simple and Easy Approach for Home Appliances Energy Consumption Prediction in Smart Homes Using Machine Learning Techniques", Journal of Applied Environmental and Biological Sciences, vol. 7, no.3, (2017), pp.108-119.

11. R. Kohavi, "The power of decision tables," in Machine Learning: ECML-95: Springer, (1995), pp. 174- 189.
12. Niembro, G. R., and R. D. Acosta, "A Methodology to Classify Residential Customers by their pattern of Consumption", In Power Engineering Society Summer Meeting, 1999. IEEE, vol. 1, (1999), pp. 226- 231.
13. Wahid, F., and D.H. Kim, "Prediction Methodology of Energy Consumption Based on Random Forest Classifier in Korean Residential Apartments", Advanced Science and Technology Letters, vol.120, (GST 2015), (2015), pp.654-687.
14. Wahid, F., and D. H. Kim, "A Prediction Approach for Demand Analysis of Energy Consumption Using K-Nearest Neighbor in Residential Buildings", International Journal of Smart Home, vol. 10, no.2, (2016), pp.97-108.
15. Wahid, F., and D. H. Kim, "An Efficient Approach for Energy Consumption Optimization and Management in Residential Building Using Artificial Bee Colony and Fuzzy Logic", Mathematical Problems in Engineering, vol. 2016, Article ID 9104735, 13 pages, (2016).
16. Li, X., Bowers, C.P. and T. Schnier, "Classification of energy consumption in buildings with outlier detection", IEEE Transactions on Industrial Electronics, vol. 57, no. 11, (2010), pp.3639-3644.
17. Liu L., H. Zhang, and Y. Zhang, "Energy consumption and potential of energy-saving of China based on industry classifications", 2011 International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE), (2011), pp. 293-299.
18. Depuru, S.S.S.R., L. Wang, V. Devabhaktuni, and P. Nelapati, "A Hybrid Neural Network Model and Encoding Technique for Enhanced Classification of Energy Consumption Data", In 2011 IEEE Power and Energy Society General Meeting, (2011), pp. 1-8. IEEE.
19. Bu, Li, D. Zhao, Y. Liu, and Q. Guan, "A hierarchical classification algorithm for evaluating energy consumption behaviors", International Joint Conference on Neural Networks (IJCNN), (2014), pp. 1461-1466. IEEE.
20. Gulbinas, R., A. Khosrowpour, and J. Taylor, "Segmentation and Classification of Commercial Building Occupants by Energy-use Efficiency and Predictability", IEEE Transactions on Smart Grid, vol.6, no.3, (2015), pp.1414-1424.
21. Wahid, F., R.Ghazali, M.Fayaz and A.S.Shah, "Statistical Features Based Approach (SFBA) for Hourly Energy Consumption Prediction Using Neural Network", International Journal of Information Technology and Computer Science (IJITCS), (2017), In Press.
22. Breiman, L., "Random forests," Machine learning, vol. 45, no.1, (2001), pp. 5-32.