Innovative Computing Review (ICR)
Volume 2 Issue 1, Spring 2022
ISSN(P): 2791-0024 ISSN(E): 2791-0032
Homepage: https://journals.umt.edu.pk/index.php/UMT-AIR

Article QR

Title:	**Energy Prediction of Appliances using Supervised ML Algorithms**
Author (s):	Robbia Gulnar Robbia[1]
Affiliation (s):	Department of Computer Science, Government College of Faisalabad (GCUF), Punjab, Pakistan.
DOI:	https://doi.org/10.32350.icr.21.05
History:	Received: April 8, 2022, Revised: May 22, 2022, Accepted: June 14, 2022
Citation:	R. Gulnar and H. Tariq, "Energy prediction of appliances using supervised ML algorithms," *UMT Artif. Intell. Rev.*, vol. 2, no. 1, pp. 00-00, 2022, doi: https://doi.org/10.32350.icr.21.05
Copyright:	© The Authors
Licensing:	This article is open access and is distributed under the terms of Creative Commons Attribution 4.0 International License
Conflict of Interest:	Author(s) declared no conflict of interest

A publication of
School of Systems and Technology
University of Management and Technology, Lahore, Pakistan

“Energy Prediction of Home Appliances using Supervised Machine Learning (ML) Algorithms”

Robbia Gulnar Robbia*

Department of Computer Science, Government College of Faisalabad (GCUF), Punjab, Pakistan.

***Abstract*-The amount of energy consumed by domestic appliances is an important area of research. Hence, the main goal of this study is to produce very precise forecasts about energy consumption by home appliances using the least amount of processing power. The algorithms used in this study for predicting energy usage/consumption included regression, K-nearest neighbor, decision trees, and random forest. These algorithms were applied on the appliances’ energy prediction dataset made available for public use at the UCI Machine Learning Repository. To compare the data sets and choose the optimal machine learning (ML) algorithm for them, root mean square error (RMSE) was computed.**

***Index Terms*- energy consumption, prediction energy utility, root mean square error (RMSE), supervised machine learning.**

I.Introduction

It is important to predict the amount of energy consumed by household electrical appliances, since improper appliance use wastes energy in the residential sector.

Hence, an accurate assessment of energy demand in the housing sector is crucial in order to determine the amount of energy that may be saved. The amount of energy saved mostly depends on the type of device being used; for example, some devices may cause an imbalance state, others may operate more slowly, and some have fixed running times. The exact energy forecasting model, from the perspective of energy providers, may assist in determining the ideal time to employ various devices to lower total carbon emissions and also to save money. By making good use of consumer business models, agents may arrange the functioning of various gadgets. The energy forecasting solutions also provide consumers with an in-depth analysis of home energy

* Corresponding author: robiagulnar@gmail.com

consumption patterns, allowing them to better manage and control their energy use and energy expenses. Reactive power management has been the subject of extensive research from the viewpoint of an industry consumer, although little research has been conducted about it from the standpoint of a household consumer. The analysis of residential energy use and customer behavior provides valuable insights that may help to develop more efficient energy consumption tactics. It is a difficult optimization problem to plan the operations of home appliances in various smart homes, since it is fundamentally a complicated nonlinear combinatorial issue [1].

Models that predict the energy consumption of home appliances have been the subject of several researches. The decision tree (DT) method, the decision table classifier (DTC), and the Bayesian network (BN) are a few examples of machine learning (ML) approaches used to provide a model for predicting the next-hour and next 24-hour energy consumption of home appliances. These methods codify expert information on energy usage and provide a suitable data structure for the regressor. They also show how challenging it is to select the optimum regression model for a particular dataset [2].

II.Literature Review

Load estimation and forecasting are key factors when it comes to efficiently distribute power and keeping reserves for the future. In order to meet the electricity demand and not to resort to load shedding, elements of load forecasting mechanisms should be integrated with the techniques utilized by the dominant power utility companies. There are many ML techniques that can be utilized to achieve the desired results. So, an in-depth analysis was carried out using multiple ML algorithms in this paper to identify the best technique. The techniques used to differentiate between all these algorithms and to list their respective advantages and disadvantages included mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). The most efficient way determined was to not only use one of these algorithms independently but to combine them in various combinations and use these combinations instead. This approach tentatively yields the most accurate results. Indeed, hybrid algorithms are the ones that work the best with the understanding that all algorithms have their specific pros and cons. Load estimation requires very frequent checks on different load types. So, different technologies have been discussed with respect to different load horizons. As far as performance and accuracy are concerned, predictive models that are a combination of more than two existing models have been proven to be the most fruitful. Support vector machine (SVM), artificial neural network (ANN), and other relevant models have achieved a well-organized power system utility along with the minimum percentage of error. The authors of the current study conducted numerous tests to determine which techniques go well together in a hybrid model and yield the most accurate results [3].

World population is increasing day by day and with it the overall electricity consumption is also increasing. In the current situation, where supply struggles to meet demand, the best course of action is to implement those techniques which may help to predict the overall electricity consumption at any given time in order to take necessary measures beforehand to meet the demand. The energy consumption trends throughout are nonlinear and remain dynamic. To predict short-term and long-term consumption with high accuracy, it is imperative to use machine learning along with distributed demand response programs. In the current paper, the techniques discussed for this purpose include logistic regression (LR), support vector machine (SVM), naïve Bayes (NB), decision tree classifier (DTC), K-nearest neighbor (KNN), and neural networks (NNs). The point is to propose a model well suited for the estimation and prediction of short term load forecasting (STLF). After extensive research and analysis, DTC was identified as the best technique. Enhanced DTC (EDTC) was proposed that utilizes integrated filter function, loss function, and gradient boosting to fine tune the already existing DTC mathematical model. The resulting EDTC algorithm yields a better forecasting result.

The forecasting and stabilizing of smart grid (SG) remains a challenge in today's landscape. ML comes into play when we discuss the resilience of the SGs. Indeed, there are different ML algorithms that have the capability to predict the future energy needs. Amongst all the suitable algorithms, selecting and adopting the best one poses a challenge. Numerous tests were conducted to select the best model and DTC outperformed all the other algorithms. EDTC was established to further amplify the predicting capabilities of DTC. EDTC was found to be superior to SVM, KNN, NN, LR, and DT, when it came to accuracy, precision loss, and ROC curve metrics [2].

It is crucial to predict and schedule the energy needs in smart buildings (SBs) and to meet them accordingly in order to deploy energy-efficient management systems In the current paper, several approaches were explored and amongst them artificial neural network (ANN) and genetic algorithms remained in focus. To get the most accurate result, ANN was implemented in a real SB testbed. The tests were conducted using poly-voltaic panel installation and SB electronic appliances and the data was collected from them. To implement ANN, the authors used CompactRIO and their model exhibited subpar prediction accuracy [4].

In the proposed paper, authors built a model for accurate prediction of the energy consumption and also worked on the scheduling process. The proposed model utilized machine learning. The research was intended as a roadmap towards a better model which can achieve accuracy greater to that of this model.

The proposed model was implemented using python in LabVIEW making a new VI. In the model, blocks containing any SB appliance could be selected any time of the day. The model functioned on the basis of ANN. The algorithm was not very accurate as the dataset fed to the model was not big enough for it to make proper computations. Hence, training and validation remained a challenge. It was determined that ANN is not the best model when it comes to prediction [4].

In this day and age, STLF is needed to fulfil the demand of power consumption. This helps to predict the pattern on which the power system operates. The methods of load forecasting employed in non-residential areas are based on customer demand or on experience and historical data. The best way to make the best prediction is to use machine learning. In the current paper, the authors sought the best ML model to generate an accurate short term algorithm for non-residential areas. For this purpose, the authors conducted experiments and rooted out the best model for industrial users. Recurrent neural network based on sliding window approach turned out to be the best model for both short-term and long-term prediction. After three months of testing, the model that gave the best results was gated recurrent unit (GRU), with long-short term memory (LSTM) as the second best. GRU minimized 5326.17 euros compared to LSTM in these three months and resulted in 5.28% MAPE. The proposed model was made to justify and evaluate the gap between evaluation matrices and the impact of forecast errors in power market. The implementation involved three-month data of different ML algorithms. GRU turned out to be the best and the authors considered the data as sufficient [5].

Considering the overall increase in population and the depletion of energy resources around the world, we need to utilize our energy resources efficiently and develop a model to accurately predict the overall energy consumption according to the given factors. To figure out the best factors utilized in the prediction process, univariate regression algorithm was employed by the authors. The algorithm predicted that the factors with the most impact were overall height, roof area, surface, and relative compaction. The models tested for the given factors were DT, RF, and K-NN. The testing of these algorithms was conducted on Orange software. After extensive testing, the algorithm that gave the most accurate prediction was RF. The forecasting error was 1.128 and 0.404 for cooling load and heating loads, respectively. Research was conducted on multiple ML algorithms and again the algorithm that yielded the best result was RF. The error rate determined from the testing turned out to be 0.404 and 1.128 for heating and cooling loads, respectively. Contributing factors were determined to make the best prediction using univariate algorithm. Height was determined as the most notable feature that contributed towards the prediction of the overall power consumption [6].

Home energy management systems (HEMS) can be further enhanced by the use of load forecasting. This can be achieved by utilizing machine learning considering the increase in the relevant data in the recent years. The current authors propose two methods for load forecasting, enhancing the traditional long short-term memory (S2S-LSTM) model. In the first method, three algorithms are applied: density based spatial clustering of applications with noise (DBSCAN), K-means, and Pearson correlation coefficient (PCC). Amongst all these techniques, PCC proved itself to be the best one. PCC was better at accommodating a large number of consumers. The second method constitutes an extension to method one and increases its performance. It utilizes NN architecture with softmax layers which are fully-connected, dropout, and stable. In LSTM, it further optimizes supervised learning which results in a more stable and accurate model for prediction. The findings were reached by conducting an 8-week long research with 2337 consumers.

In the current paper, two methods were proposed to make accurate predictions when it comes to energy consumption. These methods enhance S2S-LTSM model to make the predictions. Method one uses an amalgamation of human pattern recognitions which is extracted by three cluster analysis algorithm. Amongst all the algorithms that were employed PCC proved itself to be the best. The idea behind method one is to make a weight matrix by energy utilization habits and calculate the distance of the cluster. Three layer optimizing architecture was used to further enhance method one [7].

Electricity is a facility that is utilized daily by almost everyone and the lack of this energy will lead major disasters. We should generate only the required amount of electricity for utilization and cannot make more that required because utilizing large sums of energy is not possible. The price that is associated with the energy depends on the sources of the energy which in most cases are hydro-electric power plants, petroleum products, nuclear and wind energy plants. Under and overproduction are also the causes of the fluctuation in price, but the ones that contribute the most towards that fluctuation are metrological parameters, economics and industrial activities. That's why load estimation needs to performed on a regional level which will help in efficiently manage, scheduling and planning. All of this would result in overall low cost. In machine learning there are multiple algorithms that can be used to accurately measure and estimate the overall energy requirement. There are different algorithms that were used in the proposed paper. The different supervised learning algorithms were linear regression (LR), support vector repressor's (SVR), K-nearest neighbor (KNN), random forest (RF), and AdaBoost. The performance that was associated with all these algorithms varied with different times and data. To minimize the price per unit we utilize machine learning with correlated metrological parameters kept in consideration. The model that was the focus of the proposed paper was least cost electric load forecasting model (lcELFM). The model was implemented by minimizing root mean square error

(RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The data for testing was taken from Muzaffarabad from start of January 2014 till end of December 2015. Pakistan meteorological department provided author with meteorological time series data for the same period and time. The proposed model turned out to be the best when compared to other models.

Human energy consumption patterns change consistently with changes in habits that develop due to the change in weather. This energy is generated using sources like water, petroleum, wind, and natural energy. Consumers want the price to be less whereas the providers want the profits to be maximum [8].

AELF model was developed taking in account the weather conditions and human behaviors. This model is also used to reduce the price of electricity. The proposed paper suggests a least cost estimation model and utilized meteorological parameters driven electrical load demand of Muzaffarabad. The study suggested that meteorological parameters like temperature influence the overall consumption. This proved that factors like time and season drastically impact the consumption. The proposed model generated forecasting reduced prediction error for ELnMPFModels. There was a significant reduction in MAPE and with the implementation of the proposed model Muzaffarabad will save 0.303 million rupees daily. Although this study was conducted keeping Muzaffarabad in mind but the same model can be implemented in any city of Pakistan to get accurate estimation on load [4].

Estimation of the power consumption is a very important tasks when we are supposed to generated energy in advance or plan for the generation of power beforehand. With the implementation of smart grid, the need of energy consumption estimation is dire. Estimation of a future event is always a difficult task and to do it with high precession is an even bigger challenge. There have been many attempts at estimation of power consumption accurately but none of them very accurate. Machine learning has multiple algorithm that are well suited for this task. Machine learning has been recognized to predict failure before it even occurs. Machine learning is artificial intelligence (AI) which develop a model based studying a given data. The algorithms that were explored were artificial neural network (ANN), multiple linear regression

(MLR), adaptive neuro fuzzy interface system (ANFIS), and support vector machine (SVM). The criteria selected in proposed paper for power generation is Cyprus. Testing was done on real data accumulated over 2016 and 2017 for the motivation of using in long term and short term analysis. It was determined that the factors that affect the consumption of electricity the most are temperature, humidity, solar irradiation, population, gross national income (GNI) per capita, and the electricity price per kilowatt-hour. By doing multiple computations it was later discovered that SVM and ANN were superior to other machine learning algorithms which had fewer estimation error [9].

With smart grid implementation load estimation is more important than ever. The prediction of load in any given time is difficult considering the dynamic nature of the consumer. For prior planning of the energy consumption it is crucial to estimate it beforehand. In this study multiple methods were tested where ANN and SVM were more accurate and provided better estimation of the energy required [10].

III.Research Methods

ML algorithms, on which the data set was implemented and executed, included decision trees (DT), support vector machine (SVM), logistics regression, linear regression, K-nearest neighbor (KNN), and random forest (RF). SVM is a type of generalized linear classifier that uses supervised learning to classify data into binary categories. SVM was first presented/introduced in 1964 and it grew in popularity during the 1990s, resulting in a number of enhanced and expanded algorithms. Regression issues may also be solved with SVM. KNN may be used to study regression by getting a sample's nearest neighbors and assigning the average of these neighbors' properties to the sample in order to obtain the sample's properties. Another (improved) way is to assign different weights to the effects of neighbors at various distances from the sample, with the weight being inversely proportional to the distance. Among ML techniques, it was found that RF is faster in the training process and powerful for/more effective in solving high dimensional data and complex problems in the industry. Its performance remains stable and accurate, due to which it creates multiple decision trees and combines them to produce output. Figure 1 summarizes the proposed methodology.

IV. Experimental Setups and Results

The experimental setup included the tools needed to identify the best ML algorithm for making prediction and doing anticipatory tasks.

A. Data Set

Data set consisted of temperature (measured in Celsius) and humidity (measured in percentage). The data was collected

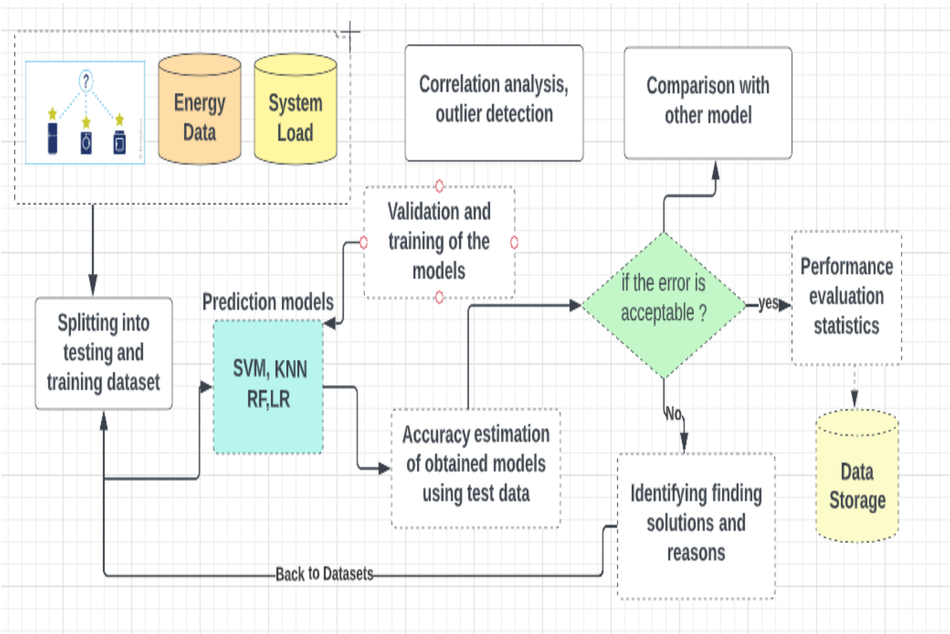

Fig. 1. Proposed method

by installing temperature sensor and humidity sensor in all the rooms of a low energy house. The values on the temperature sensor and humidity sensor were monitored with the help of ZigBee wireless sensor network. Each sensor transmitted the data (condition of temperature and humidity) for 3.3 mins, which was averaged for a 10 mins period. The data was monitored for every 10 mins and logged in. The details about the outside weather, pressure, humidity, wind speed, and visibility were taken from the weather station. This data was linked with the experimental data of temperature and humidity of each room with the help of date and time columns. The source of energy prediction for appliances data set is UCI Machine Learning Repository. Table 1 summarizes the characteristics of the dataset.

From a total of 19,735 data samples, 15,788 samples (80%) were randomly assigned to the training set and 3,947 samples (20%) to the testing set. Table 2 gives the data set description.

Table I

Characteristics of Dataset

Characteristic	Values
No. of data sample	19,735
No. of features	29
Sampling time/rate	10 min

B. Results

All prediction models are regression models. Root mean square error (RMSE), mean absolute error (MAE), mean square error (MSE), and decision coefficient (R2) are all regularly used metrics for assessing regression models. In this experiment, the RMSE assessment indicator was employed. It was used to calculate the difference between the observed value and true value.

$$\text{RMS}=\sqrt{1/m\sum_{i=1}^{m}(Yi-Yi')}$$

Here, *Yi* is the real value of the data at time *i* and *Yi'* is the predicted value of the data at time *i*.

Table 3 shows the performance evaluation of the prediction models used to measure the energy consumption of household appliances. RMSE of linear regression is 0.046 (which is very low). It shows the better performance of this model as compared to other models. RF has RMSE of 0.047 and it also shows better performance than SVM and logistic regression.

Table III

Model Performance

Models	Accuracy	RMSE
Linear Regression	0.953	0.046
Random Forest	0.939	0.047
Logistic Regression	0.732	11.71
SVM	0.740	10.37

Table II

Dataset Variables and Description

Feature #	Variable	Description	Unit
1	T1	Temperature in kitchen	°C
2	T2	Temperature in living room	°C
3	T3	Temperature in laundry room	°C
4	T4	Temperature in office room	°C
5	T5	Temperature in bathroom	°C
6	T6	Temperature outside the building	°C
7	T7	Temperature in ironing room	°C
8	T8	Temperature in teenager room	°C
9	T9	Temperature in parents' room	°C
10	R1	Humidity in kitchen	%
11	R2	Humidity in living room	%
12	R3	Humidity in laundry room	%
13	R4	Humidity in office room	%
14	R5	Humidity in bathroom	%
15	R6	Humidity outside the building	%
16	R7	Humidity in ironing room	%
17	R8	Humidity in teenager room	%
18	R9	Humidity in parents' room	%
19	L	Light energy consumption	Wh
20	RO	Humidity outside the airport	%
21	Td	Dew point temperature	°C
22	V	Visibility	km
23	W	Wind speed	m/s
24	TO	Temperature outside the airport	°C
25	rv1	Random Variable1	\
26	rv2	Random Variable2	\
27	P	Pressure	mmHG

Fig 2. Accuracy of Models

Fig 3. Root Mean Square Error of Models

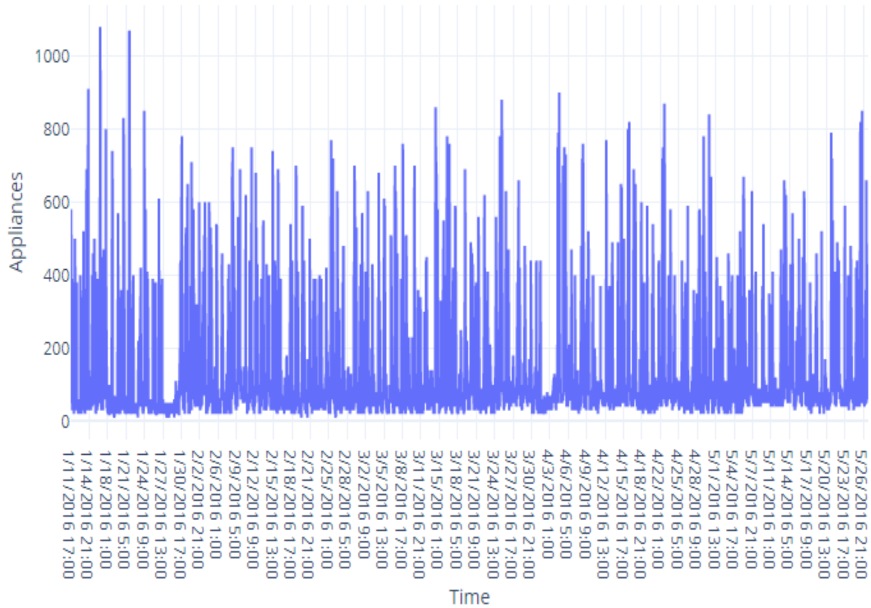

Fig 4. Energy consumption by home appliances in the given period

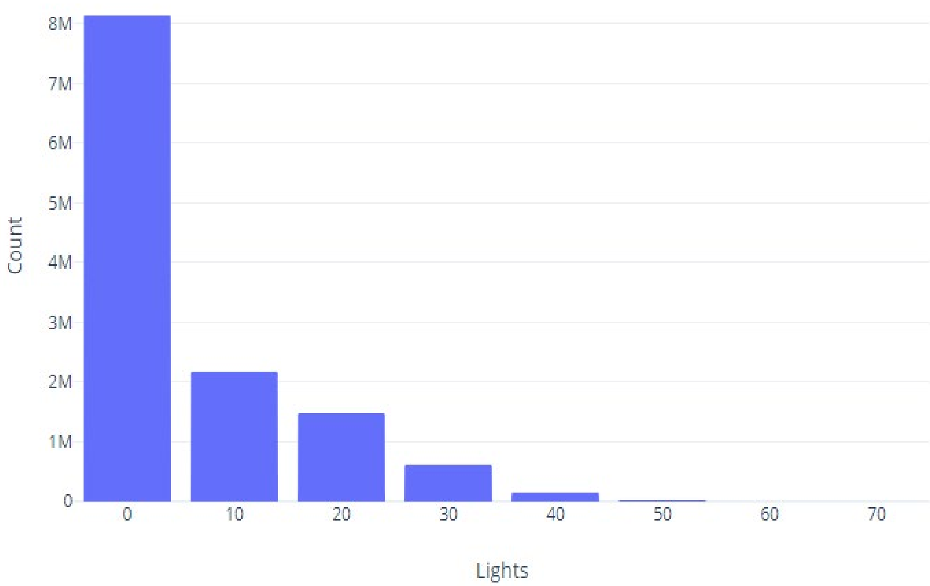

Fig 5. Energy Consumption of Lights

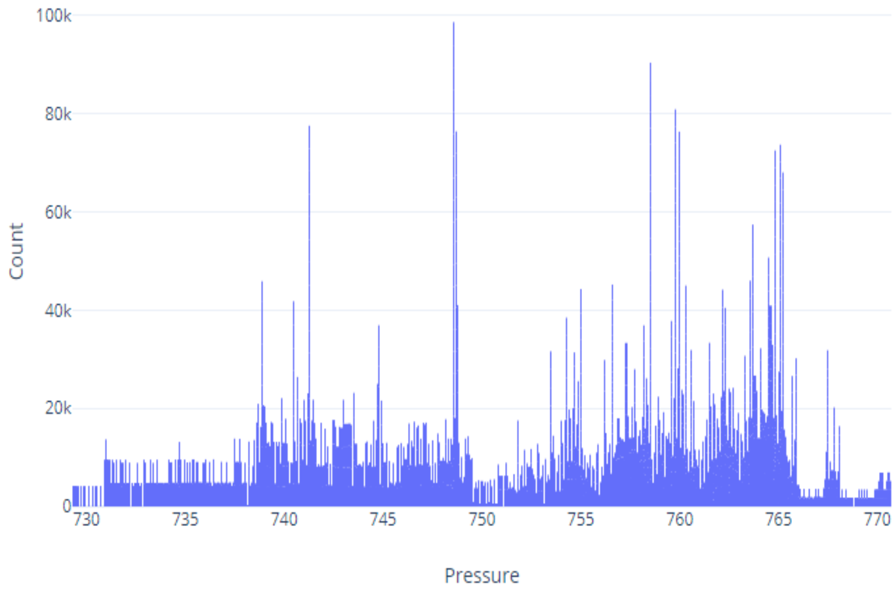

Fig 6. Pressure and Humidity during Energy Consumption

Fig 7. Graph Depicts Visibility (Scatter Plot)

V. Conclusion

Prediction models for energy usage by home appliances based on SVM, KNN, RF, linear regression, and logistic regression were investigated. Firstly, the authors reviewed the data pretreatment to remove certain features from the filtered data and to normalize it. Secondly, the grid search technique was utilized to find the best parameters for the model and models based on several ML algorithms were created. Finally, each model's prediction performance was tested and compared. The results revealed that linear regression and RF obtained good results in both the training and testing data sets, with the best prediction performance among the four prediction models developed using the classic ML approach. In the testing set, KNN, RF, and SVM all performed equally well, however, SVM performed the poorest in the training data set.

References

[1] S. Hu, D. Yan, S. Guo, Y. Cui, and B. Dong, "A survey on energy consumption and energy usage behavior of households and residential building in urban China," *Energy Build.*, vol. 148, pp. 366–378, Aug. 2017, doi: https://doi.org/10.1016/j.enbuild.2017.03.064

[2] T. Alquthami, M. Zulfiqar, M. Kamran, A. H. Milyani, and M. B. Rasheed, "A performance comparison of machine learning algorithms for load forecasting in smart grid," *IEEE Access*, vol. 10, pp. 48419-48433, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3171270

[3] A. A. Mamun, M. Sohel, N. Mohammad, M. S. Haque Sunny, D. R. Dipta, and E. Hossain, "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," *IEEE Access*, vol. 8, pp. 134911-134939, 2020, doi: https://doi.org/10.1109/ACCESS.2020.3010702

[4] J. Zhua, Y. Lin, W. Lei, Y. Liu, and M. Tao, "Optimal household appliances scheduling of multiple smart homes using an improved cooperative algorithm," *Energy*, vol. 171, pp. 944-955, Mar. 2019, doi:

https://doi.org/10.1016/j.energy.2019.01.025

[5] S. Ungureanu, V. Topa, and A. C. Cziker, "Analysis for non-residential short-term load forecasting using machine learning and statistical methods with financial impact on the power market. *Energies*, vol. 14, no. 21, Art. no. 6966, doi: https://doi.org/10.3390/en14216966

[6] S. Hosseini and R. H. Fard, "Machine learning algorithms for predicting electricity consumption of buildings," *Wirel. Pers. Commun.*, vol. 121, no. 4, pp. 3329-3341, Aug. 2021, doi: https://doi.org/10.1007/s11277-021-08879-1

[7] L. Fan, J. Li, and X.-P. Zhang, "Machine learning load prediction methods for home energy management systems based on human behavior patterns recognition," *CSEE J. Power Energy Syst.*, vol. 6, no. 3, Sep. 2020, doi: https://doi.org/10.17775/CSEEJPES.2018.01130

[8] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Machine learning for energy performance prediction at the design stage of buildings. *Energy Sustain. Dev.*, vol. 66, pp. 12-25, Feb. 2022, doi: https://doi.org/10.1016/j.esd.2021.11.002

[9] D. Solyali, "A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus. *Sustainability*, vol. 12, no. 9, Art. no. 3612, 2020, doi: https://doi.org/10.3390/su12093612

[10] K. Thiyagarajan and R. S. Kumar, "Real time energy management and load forecasting in smart grid using compact RIO," *Procedia Comput. Sci.*, vol. 85, pp. 656–66, 2016, doi: https://doi.org/10.1016/j.procs.2016.05.250