

# Residential Electricity Demand Prediction using Machine Learning

Manpreet Kaur, Shalini Panwar, Ayush Joshi, and Kapil Gupta

National Institute of Technology, Kurukshetra Haryana, India

## Abstract

This paper presents an analysis of the usage of electric power in the residential sector and predicting the demand for power consumption of the next day and aims to improve the prediction accuracy and find the best model and try to reduce the cost of overall power consumption in a building. Consumption of electric power can be broadly divided into two categories i.e. commercial and residential sectors. This procedure consists of three steps i.e. feature extraction, normalization, and validating. Heavy fluctuation may arise in the residential sector may cause damage to electrical appliances. To match the demand of customers and generation of power at generating unit prediction is necessary. A variable power pattern may cause stress at power grid. Prediction of electric consumption is required in prior so that load at the power grid could be balanced. To meet the requirement of demand, appliances can be swapped from peak hours to an off-peak hour and also reduce the cost at the customers' side. The performance of different models can be compared by using different evaluating indices are: Coefficient of determination ( $R^2$ ), mean absolute error (MAE), mean squared error (MSE). Out of Linear Regression, Lasso Regression, Ridge Regression, Elastic Net Regression, Random Forest Regression, Extra Trees, Support Vector Regression, Decision tree. Lasso Regression and Support Vector Machine outperforms with an accuracy of 99.99% and 99.89% with 0.01% and %0.11 % mean squared error respectively.

## Keywords 1

Prediction, Demand, Consumption, Residential, Generation, Power, Dynamic, Regression, Learning, Temperature, Building.

## 1. Introduction

As we know electric power plays a vital role in today's era. To generate the power accordingly so that it can fulfill the demand of customer is not an easy task as natural sources of generation of power is extinguishing day by day. The total generation of power is distributed among various sectors according to their requirement. Sectors can be residential and commercial (offices, factories). For now, we are just focusing on the residential sector. It is not necessary that all the electrical appliances will consume the same amount of power. Appliances may have to undergoes various

issues i.e. fluctuation. When we receive more or less power than its requirement which can degrade the performance of the appliances. The consumption of power by appliances is dependent on various input parameters i.e. type of day [12]. If it's summer season, the Air-conditioners appliance will consume more power. In the winter season, heating appliances will consume more power whereas, in the rainy season, lighting will consume more power. The total cost of power in the residential sector also depends on the type of the hour i.e. peak hour, off-peak hour, mid-peak hour [11] in addition to the amount of power consumed. During peak hours, load at a power grid, and cost of power

ISIC'21: International Semantic Intelligence Conference, February 25-27, 2021, New Delhi, India  
 EMAIL: [mannri346@gmail.com](mailto:mannri346@gmail.com) (M. Kaur),  
[panwarshalini40@gmail.com](mailto:panwarshalini40@gmail.com) (S. Panwar),  
[ayushjoshi75.aj@gmail.com](mailto:ayushjoshi75.aj@gmail.com) (A. Joshi);



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

per unit hour is more as compared to an off-peak hour, load at a power grid, and cost of per unit hour is least. To handle this issue, appliances can be swapped from peak hour to off-peak hour or we can say appliances can be scheduled in order to manage the overall load of the power grid and total cost of customer. To schedule the appliances, the power generating company and the customers make a deal in order to compensate either by a price based [1][2][3][4] or incentive-based [4][5][6] demand response type. To meet the availability of electric power and its demand we have to make predictions. Wrong or no prediction may lead to a violation of the agreement of service level. Consumption of electric power is directly dependent on climatic conditions. Either if we make a prediction still, we can't guarantee the weather of the next second as the weather changes drastically. If the customer knows they need consumption of power, then they can make long-term plans. Even dynamic pricing may help to know the price of the power in the next hour. The biggest challenge in the prediction of building like indoor temperature, outdoor temperature because of climate change. It is very difficult to predict the effect on other parameters by increasing or decreasing its value during the construction of the building. As it was already discussed that it is not possible to include all the key features or we can say wide range of building features and general principle to predict the demand of a power in residential sector. But in this, author stated one model i.e. dynamic high resolution demand side management which combine all the building features and general principle [15]. And also analysed that maximum of the energy consumption is consumed by lights and building integrated photovoltaic [16]. Different machine learning algorithms i.e. Linear Regression, Elastic Net, Decision Tree, Random Forest, SVM, etc. out of all these models Linear Regression is the benchmark model for the prediction, and the performance of different models can be compared by using different evaluating indices are: Coefficient of determination ( $R^2$ ) [7], mean absolute error (MAE) [7], mean squared error (MSE).

This paper is organized as follows: Section II related work of machine learning algorithms in which 8 Regression algorithms used for prediction are briefly explained. In section III, 4 evaluation indices are used to evaluate the performance of prediction models. In section

IV, experimental settings of the models which explain the dataset and how the parameters of models are tuned. In section V, experimental observation and results. Section VI, finally concludes the paper.

## 2. Related Work

### 2.1 Linear Regression

Linear Regression is used to find the relationship between predictor or independent variable and target or dependent variable. If one variable is expressed accurately to another variable then it is known as deterministic. The basic idea of the linear regression model is to derive the best fit line, also known as the regression line. Sum of the distance between the points on the graph and regressor line are the total prediction error of all the data points. Smaller the error, the better the result and vice versa [8].

Linear regression Equation:

$$Y(pred) = b_0 + b_1 * x \quad (1)$$

$b_0$  is intercepting whereas  $b_1$  is the slope of the regression line. In order to get the minimum error, the value of  $b_0$  and  $b_1$  should be minimum.

The error between the predicted value and the actual value can be calculated as:

$$Error = \sum (predicted_{value} - actual_{value}) \quad (2)$$

Exploring the value of 'b1':

If  $b_1 < 0$  then it will have a negative relationship, which means a decrease in target value with an increase in predictor value. If  $b_1 > 0$  then it will have a positive relationship, which means the value of the target will increase with the increase in the value of predictors.

Exploring the value of  $b_0$ : If the predictor is 0 then the equation will be meaningless and of no use.

In Figure 1, it shows the graph of predicted value of power in a residential sector.

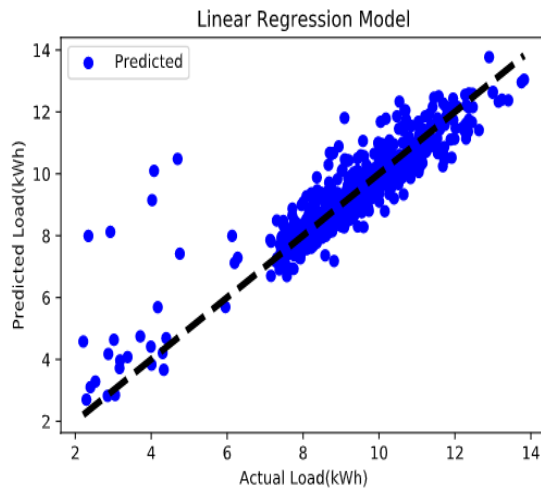


Figure 1: Linear Regression

## 2.2 Lasso Regression

It stands for Least Absolute Shrinkage and Selection Operator. From its full form, it is clear that it uses shrinkage and it is a type of Linear Regression. Here, shrink means that values of the dataset will be shrunk towards the central point, say similar to that of Mean. The performance of this model is good when the dataset contains multicollinearity.

Lasso Regression undergoes L1 Regularization means it is the summation of the absolute value of the magnitude of the coefficient. Here, a few of the coefficients can be zero and that values can be eliminated from the dataset. Larger penalties will result in the coefficient values near zero whereas smaller penalties will result in the coefficient values far away from zero. The aim of this algorithm is to minimize the error

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} * \beta_j)^2 + \lambda \quad (3)$$

$$\sum_{j=1}^p |\beta_j|$$

If  $\lambda = 0$  means there is an absence of regularization and thus we get Ordinary Least Squares solution. When  $\lambda \rightarrow \text{INF}$ , then coefficients will lead to 0 and the model left out be a constant function To tune the parameters,  $\lambda$  is the amount of shrinkage when  $\lambda = 0$ , parameters will not be eliminated. When  $\lambda$  increases, bias also increases whereas when  $\lambda$  decreases, variance also decreases. Bias and variance are inversely proportional to each

other. L1 regularization will results in a sparse model.

It is a challenging task to select one variable as the predictor which particulates suite the property of Lasso Regression. The selection can be done haphazardly but it can result in a very bad decision means a very time-consuming process.

## 2.3 Ridge Regression

If an overfitting or underfitting type of problem arises then there are chances that it works as linear regression. Ridge Regression is a method to create a parsimonious model i.e. when the number of predictor values is more than the number of observations means when there is a correlation between predictor values i.e. the dataset has multicollinearity.

Tikhivov's method has a larger set as compared to the parsimonious model but it is similar to ridge regression. If a dataset contains a noise i.e. statistical noise still this model can produce the solution.

Ridge regression undergoes L2 regularization. Also known as the L2 penalty. coefficients of data values are shrunk by the same factor and none of the value is eliminated. Unlike L1 regularization, L2 will not result in a sparse model.

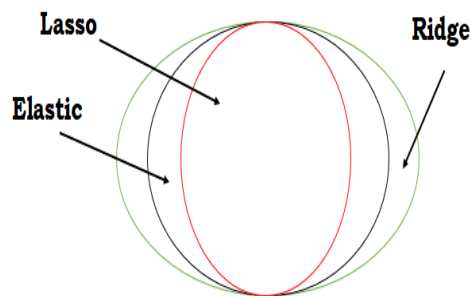
$$\sum_{i=1}^n (y_i - y_i')^2 = \sum_{i=1}^n (y_i - \sum_{j=0}^p w_j * x_{ij})^2 + \lambda * \sum_{j=0}^p w_j^2 \quad (4)$$

To strengthen the term of penalty, we have to tune the parameter i.e.  $\lambda$  When  $\lambda$  is 0, least squares and ridge regression are equal. When  $\lambda$  is  $\infty$ , all coefficient will be zero. The overall penalty will range from 0 to  $\infty$  Overall, Least Square uses the following equation:

$$B' = (X'X)^{-1} X'Y \quad (5)$$

Here, X is a scaled and centered matrix. When columns of the X matrix have high multicollinearity then the cross product of  $(X'X)$  matrix will be singular or nearly Singular. Including ridge parameter (k) to the above equation, then the new equation will be

$$B' = (X'X + kI)^{-1} X'Y \quad (6)$$



**Figure 2:** Lasso Regression, Ridge Regression, Elastic Net Regression

## 2.4 Elastic Net Regression

It is a technique that uses properties of the L1 penalty (Lasso Regression) and L2 penalty (Ridge Regression). To improve the regularization, we combine both lasso and ridge regression. It is a 2-step process i.e. in the first step it finds the coefficient of ridge regression by selecting group feature and in the second step, it performs lasso sort of the coefficients of shrinking by performing feature selection. The objective of this model is at minimizing by using the following equation:

$$L_{\text{enet}}(\beta) = \frac{\sum_{i=1}^n (y_i - x_i\beta)^2}{2n} + \lambda \left( \frac{1-\alpha}{2} \right) \quad (7)$$

$$\sum_{j=1}^m \beta_j * \beta_j + \alpha \sum_{j=1}^m |\beta_j|$$

Here,  $\alpha$  is the mixing parameter i.e.  $\alpha = 1$  reduces the function to lasso regression whereas  $\alpha = 0$  reduces the function to ridge regression. Parameter  $\lambda$  is highly dependent on the  $\alpha$  parameter. It has better predictive potential than lasso regression.

One of the biggest disadvantages of Elastic Net Regression is that it may or may not remove all the irrelevant coefficients.

In Figure 2, it shows the relationship between Lasso Regression, Ridge Regression and ElasticNet Regression.

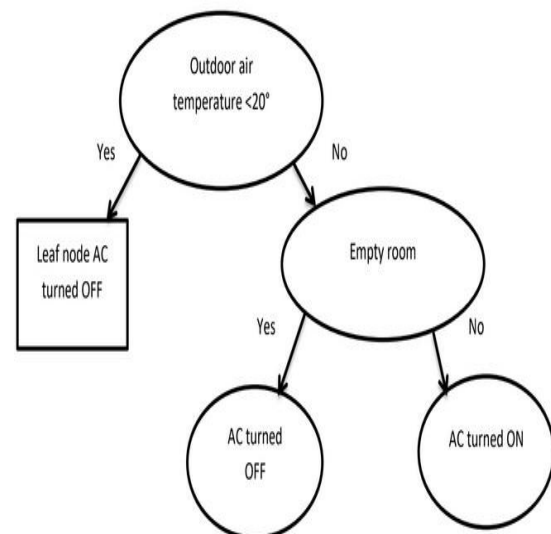
## 2.5 Decision Tree

It is a supervised machine learning algorithm. From the name, it defines that it is a decision-making tool and it uses flowchart like tree structure [8]. It supports both the

continuous and discrete output values. Here, continuous output example is to predict the required power of the building where our ultimate goal is to reduce the overall cost of the power whereas discrete output value means to predict the rain on a particular day that whether it rains or not.

Decision Nodes are known as conditions of a flowchart whereas terminals are known as results of a flowchart. The root node is called as best predictor node as this node is the topmost decision node. Every machine learning algorithm model has its advantages and disadvantages but the advantage of the decision tree is that it is a very good model at handling the tabular data with categorical features with lesser than hundreds of categories and numerical data.

The decision tree can capture the non-linear interaction between the predictors and the target value. Suppose, target variable is air-conditioner and predictor variable is room occupancy (empty or not) and outdoor air-temperature ( $\leq 26^\circ \text{C}$ ) see figure 3



**Figure 3:** Decision Tree

## 2.6 Random forest

It is a supervised machine learning algorithm. The random forest can perform both classification and regression problems. The random forest contains multiple decision trees and the output of this is not only dependent on one decision tree but every single decision tree. Every tree is independent, none of any tree has interaction with each other while building the

model. All these trees run parallelly but independently. Every tree performs its prediction and these predictions are aggregated and perform arithmetic mean on that to produce a single final result. It can be formulated as:

$$g(x) = f_0(x) + f_1(x) + f_2(x) + \dots + f_n(x) \quad (8)$$

Here,  $g(x)$  is a single final result whereas  $f_i(x)$  is a decision tree.

Each Decision tree can be drawn using a random sample from the original dataset by splitting it and add randomness to it to prevent it from overfitting. Random forest is one of the highly accurate models which can handle thousands of predictors without the deletion of any variable.

From Figure 4, it is clear that Random Forest is multiple Decision Trees with multiple features.

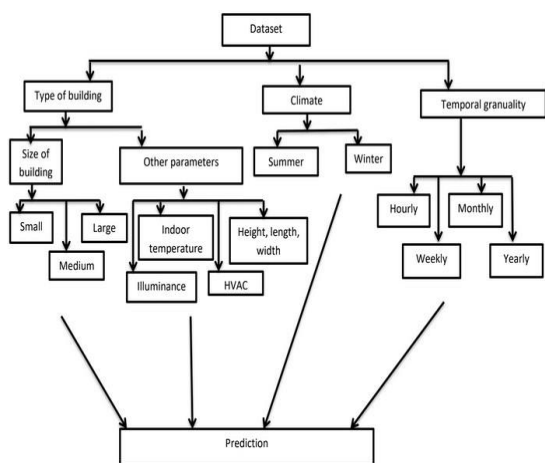


Figure 4: Random Forest

### 2.7 Extra Trees

It is also known as Extremely Randomized Trees. Unlike, Random Forest and Decision Tree, Extra Trees makes the next best split from the uniform random splits from the subsets of features and can't be substituted with another sample. Extra Trees creates a greater number of unpruned Decision Trees. Unlike Random Forest, it makes random split. In addition to the optimization of algorithms, it also adds randomization. This model is faster than other models. It takes less time to compute as it doesn't have to select the optimal split but a random split.

### 2.8 Support Vector Regression

This algorithm is one of the most popular algorithms for regression problems. Basically [8], it draws a boundary line or straight line so that n-dimensional space can be segregated into classes. The Boundary line is drawn in such a way that it can cover maximum data points between them. This boundary line is known as a hyperplane. There are two types of SVR:

**Linear SVR:** This type of data is known as linearly separable data. It draws a single straight line to differentiate two classes.

**Non-Linear SVR:** This type of data is known as non-linear separable data. It is not possible to segregate data into classes by just one single line.

This linear and non-Linear data is handled by the SVR kernel. Kernel Helps to find and draw the hyperplane without increasing the cost in n-dimensional space. Sometimes it is not possible to find the hyperplane in n-dimensional space. So, we draw n+1 dimensional space. The value of kernel can be poly, RBF, sigmoid, gaussian for non-linear datasets whereas for linear dataset value should be linear kernel only to solve the problems.

Cross-Validation is also one of the techniques which can be used in Support Vector Regression from the training purpose of the model and then evaluate the model. It is failed to generalize the pattern of the dataset but can detect the fitting whereas cross-validation is used to find the most accurate value but it may fail to enhance the accuracy.

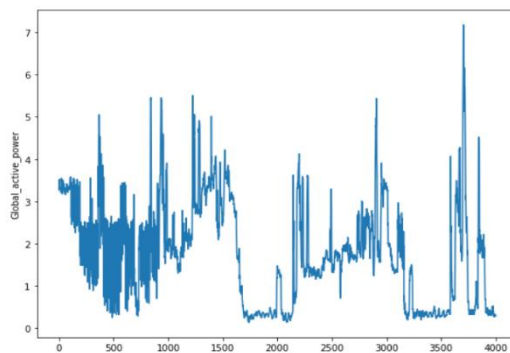


Figure 5: Support Vector Regression

Figure 5 shows the result of Support Vector Regression.

### 3. Evaluation indices

Evaluation indices are considered to evaluate the model by various authors. The performance can be checked by finding the accuracy and error of the models. The lesser the error, the more the accuracy is better than the model.

$R^2$  [7][13] (Coefficient of Determination) and RMSE [13][14] (Root Mean Squared Error) are the two methods for the model optimization whereas MAE [7][14] (Mean Absolute Error) and MSE (Mean Squared Error) two methods to evaluate the model means to check the error and tries to reduce it.  $R^2$  its value varies between 0 to 1 defining the accuracy of the model.

$$R^2 = \frac{res}{tot} \quad (9)$$

Here, the res is the sum of the square of the residual error whereas tot is the total sum of the error. If  $R^2 > 0$  it means the result is accurate, if  $R^2$  means the same result and  $R^2$  is ambiguous results.

$$res = \sum_{i=1}^n (y_i - y_i')^2 \quad (9 \text{ i})$$

$$tot = \sum_{i=1}^n (y_i - y_{mean}')^2 \quad (9 \text{ ii})$$

RMSE a metric and it is dependable on the scale.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i')^2}{n}} \quad (10)$$

MAE is also dependable on the scale. Basically, it finds the absolute of all the data points either if they have a negative error or positive error. None of any error cancels out the effect of each other

$$MAE = \frac{\sum_{i=1}^n |y_i - y_i'|}{n} \quad (11)$$

MSE is similar to that of MAE but instead of absolute the values, it squares the values then finds the Mean error.

$$MSE = \frac{\sum_{i=1}^n (y_i - y_i') * (y_i - y_i')}{n} \quad (12)$$

In the above equations,  $y_i$  is the predicted value whereas  $y_i'$  is the actual values and n is the number of data points on the graph

### 4. System Architecture

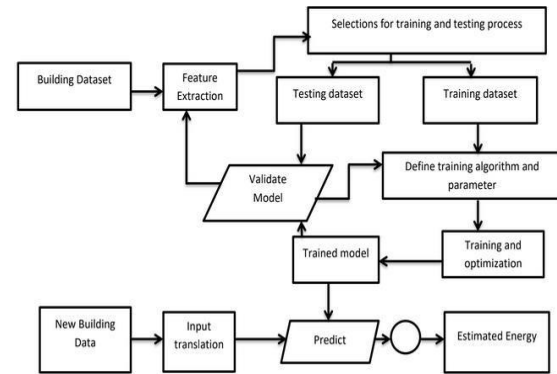


Figure 6: System Architecture

From the System Architecture figure i.e. Figure 6, The prediction of power input parameters will be of that type only like the type of a day i.e. summer, winter, rainy. The material used for construction is used i.e. if the material is insulation then it would be best. Dataset is a daily, hourly basis, yearly basis, etc. Other details of buildings can also be included i.e. height, width, illumination, occupancy, etc. Even dataset can be of 3 types i.e. Real data, Simulated data, Sensor-based data [8].

After analyzing the dataset, it undergoes the feature extraction phase, in which filtration of the dataset is done i.e. unusual data and noise is discarded and only useful data is left behind and thus undergoes transformation process i.e. dataset is transformed according to the requirement of the algorithm and after that size of the dataset is decreased to increase the performance and this process is known as reduction of dataset.

After feature extraction, transformation and reduction, the entire dataset is divided into training and testing dataset and there is a training and testing phase of the model. In the training of a model first we have to select an appropriate algorithm for the prediction and thus training can be done in two ways i.e. First principle approach i.e. the prediction of power is done based on the current situation rather than observing the history or Data-Driven approach i.e. want to give detailed information about building and thus results are validated and accuracy is measured. If it ends with good accuracy then our algorithm is ready for an unknown dataset of the building and prediction there demand of power. Thus, results are compared based on evaluation metrics and

declare one model as the best model with greater accuracy and minimum error.

## 5. Experimental Settings

Using various libraries of python such as pandas, scipy we can carry out the analysis. For the basic implementation of mining of data or ML we are analysing the library "SKlearn". Sklearn is a module of python language which integrates all the ML algorithms in the world of different python libraries such as NumPy, sciPy, Matplotlib. We are trying to use this as the study shows that this gives efficient and simple solutions. Now training and testing can be done in different models and then compared from each other in-order to get the better outcomes. From the review of different authors and according to our online study we can say that ensemble models of machine learning give better performance than others. For the time being the study about different models is being done and a dataset is collected. Also, the gradient boosting may give us the accurate results. The gradient boosting is the algorithm which also trains various models in gradual, additive and sequential manner. Since this algorithm is prone to overfitting therefore it uses hyperparameter tuning. This analysis totally depends on how better accuracy we are demanding according to the dataset available. On the other hand, Random forests also prove to be good for the efficient results. It is the algorithm which uses a special process known as "Early Stopping" in which training stops once the performance on testing data stops improving further. This is an optimized technique. Also, this avoids overfitting. Therefore, this also can be the model for our project which is to be applied and analysed. And this is beneficial for categorization as well as regression problem. Also, this can be modelled for categorical values. After the search results, this model can also be compared for the study purpose gathering more dataset and then visualizing for more accuracy and better performance for achieving the results. The search result done after training can be gathered in a document and the result could be concluded.

Experiments should be done to declare one model as the best model and for experiments, we need a dataset. Generally, the dataset can vary from 2 weeks [9] to 4 years [10]. So, the

dataset is taken from Kaggle and we uploaded on GitHub, link: <https://raw.githubusercontent.com/navkapil/googlecolab/master/pwrpred.csv>

The dataset consists of 1048576 rows and 9 columns. It contains per minute data of the day for approx. 2 years from 16-12-2006 17:24:00 to 13-12-2008 21:38:00. 9 columns of the dataset are DateTime, global active power, global reactive power, Voltage, Global intensity, sub-metering 1, sub-metering 2, sub-metering 3, and sub-metering 4. Out of all these parameters, global active power is taken as output, and the rest all other as input parameters. This dataset is divided into training and testing with a percentage of 90% and 10% respectively. After this division, it undergoes normalization so that all the parameters lie in the same range. Models for the experiments are Linear Regression, Decision tree, Random Forest, Extra Trees, Lasso Regression, Ridge Regression, Elastic Net Regression, Support Vector Regression. The result of the dataset also depends on the type of the dataset i.e. whether it is a linear or Non-Linear Dataset. All the parameters of a model have Default values. The performance of SVR is affected if data is linear and we set the value of kernel as non-linear (RBF, poly, gaussian) and vice-versa. The kernel we have taken is RBF. We checked values of C and gamma varies from .01 to 100 and 0.01 to 10 by 10 units respectively and we got good results at C=100 and gamma = 0.1 and degree as 3 (default). Linear Regression is the baseline model for this project because it gives very good results by taking default values of all the parameters. Decision Tree, Random Forest, and Extra Trees are somewhat similar models. Indecision tree default value of random\_state is None but we give it as 42 as this parameter dominates the randomness of the estimator, in random forest n\_estimators tells the number of trees to be formed and we checked the performance of the model by varying its value from 10 to 100 and got a better result for value as 50 whereas the value of n\_estimators in Extra trees is 150 but its default value is 100. If we set the default value of alpha in Lasso Regression then it works as a Linear Regression but it is not advised. For better performance, we set its value as 0.01 whereas in Ridge Regression if we set the default value of alpha then it works as a Logistic Regression so it would be better if we tuned them with value as 0.1. Basically, alpha regularization improves the problem

conditioning and hence lower down the variance of the estimators. For Elastic Net Regression, we tuned two parameters that are alpha and l1\_ratio. For alpha = 0 it is solved by Linear regression. If l1\_ratio = 0 then penalty is l2 i.e. Ridge Regression and if l1\_ratio = 1 then penalty is l1 i.e. Lasso Regression.  $0 < l1\_ratio < 1$ , is the combination of l1 and l2 penalty. Rest all the parameters of the models will be the default value.

## 6. Experiment Result

It shows the results of different models using 4 evaluation indices i.e., RMSE,

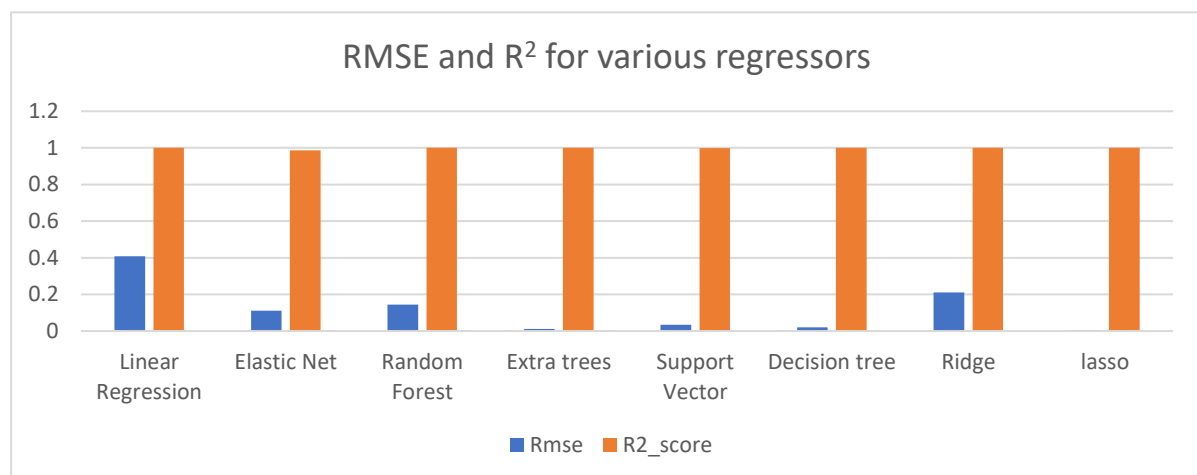
R2\_score, MSE, MAE for the prediction of next-day power consumption.

From the table i.e. Table 1, it is clear that Support Vector Machine and Lasso Regression give better accuracy with the minimum error where Linear Regression is the worst performer out of these 8 models that is why Linear Regression is taken as Baseline Model and Support Vector Machine as a benchmark system.

Figure (7) (8) shows that Elastic Net, Random Forest, Extra Trees, Support Vector Regression, Lasso all have nearly the same results but Support Vector Regressor and Lasso give good results.

**Table 1:**  
Summary of results of models

| Models                   | RMSE    | R2_score | MSE     | MAE     |
|--------------------------|---------|----------|---------|---------|
| <b>Linear Regression</b> | 0.4088  | 0.9999   | 0.1671  | 0.235   |
| <b>Elastic Net</b>       | 0.11204 | 0.9862   | 0.01255 | 0.0916  |
| <b>Random Forest</b>     | 0.1451  | 0.9998   | 0.02    | 0.0345  |
| <b>Extra trees</b>       | 0.01088 | 0.99988  | 0.012   | 0.0248  |
| <b>Support Vector</b>    | 0.03404 | 0.9989   | 0.0011  | 0.02706 |
| <b>Decision tree</b>     | 0.01952 | 0.9999   | 0.3026  | 0.0413  |
| <b>Ridge</b>             | 0.2112  | 0.9999   | 0.1238  | 0.7048  |
| <b>lasso</b>             | 0.04626 | 0.99990  | 0.0001  | 0.01119 |



**Figure 7:** RMSE and R<sup>2</sup>



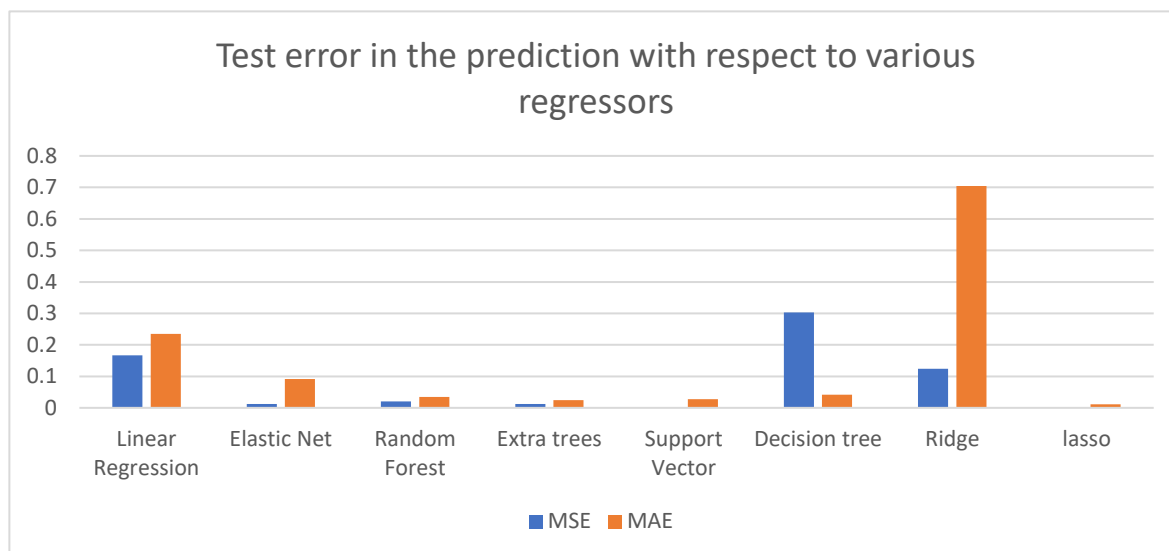


Figure 8: MSE and MAE

## 7. Conclusion

This paper focuses on the implementation of various machine learning algorithms for predicting the power of buildings. It is not necessary that for a particular dataset it will always show a good result, sometime it may show the uncertain results as every model have their pros and cons. By varying one or more parameters of the building what would be its effect on the other parameters, it would be very difficult to predict as we can't set all the parameters according to our requirements. Because of prediction, it would be very easy to make long term plans. Weather data plays a vital role in the prediction of the power of a building. According to our analysis of the result of the model's Support vector regression, lasso regression is the best model. Even to increase the accuracy of these models we can use Long Short Term Memory (LSTM) as it is very robust deep learning algorithm for prediction of time based forecasting and has potential to give accurate prediction results or hybrid approach.

## 8. References

- [1] Setlhaolo, D., Xia, X., & Zhang, J. (2014). Optimal scheduling of household appliances for demand response. *Electric Power Systems Research*, 116, 24-28.
- [2] Gayatri, P., Sukumar, G. D., & Jithendranah, J. (2015, December). Effect of load change on source parameters in power system. In 2015 Conference on Power, Control, Communication and Computational Technologies for Sustainable Growth (PCCCTSG) (pp. 178-182). IEEE.
- [3] Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, 11921205.
- [4] Gomes, Á., Antunes, C. H., & Oliveira, E. (2011). Direct load control in the perspective of an electricity retailer—a multi-objective evolutionary approach. In *Soft Computing in Industrial Applications* (pp. 13-26). Springer, Berlin, Heidelberg.
- [5] Babar, M., Ahamed, T. I., AlAmmar, E. A., & Shah, A. (2013). A novel algorithm for demand reduction bid based incentive program in direct load control. *Energy Procedia*, 42, 607613.
- [6] Liu, D., & Chen, Q. (2013, June). Prediction of building lighting energy consumption based on support vector regression. In 2013 9th Asian Control Conference (ASCC) (pp. 1-5). IEEE.
- [7] Muralitharan, K., Sakthivel, R., & Shi, Y. (2016). Multiobjective optimization technique for demand side management with load balancing approach in smart grid. *Neurocomputing*, 177, 110-119.
- [8] Amasyali, K., & El-Gohary, N. (2016). Building lighting energy consumption prediction for supporting energy data analytics. *Procedia Engineering*, 145, 511-

- 517.
- [9] Liu, D., & Chen, Q. (2013, June). Prediction of building lighting energy consumption based on support vector regression. In 2013 9th Asian Control Conference (ASCC) (pp. 1-5). IEEE
- [10] Dagnely, P., Ruelle, T., Tourwé, T., Tsiorkova, E., & Verhelst, C. (2015, September). Predicting hourly energy consumption. can regression modeling improve on an autoregressive baseline?. In International Workshop on Data Analytics for Renewable Energy Integration (pp. 105-122). Springer, Cham.
- [11] Naji, S., Çelik, O. C., Alengaram, U. J., Jumaat, M. Z., & Shamshirband, S. (2014). Structure, energy and cost efficiency evaluation of three different lightweight construction systems used in low-rise residential buildings. *Energy and buildings*, 84, 727-739.
- [12] Ali, S., Ahmad, R., & Kim, D. (2012, December). A study of pricing policy for demand response of home appliances in smart grid based on M2M. In 2012 10th International Conference on Frontiers of Information Technology (pp. 231-236). IEEE.
- [13] Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. *European journal of operational research*, 199(3), 902-907.
- [14] Gonzalez-Romera, E., Jaramillo-Moran, M. A., & Carmona-Fernandez, D. (2006). Monthly electric energy demand forecasting based on trend extraction. *IEEE Transactions on power systems*, 21(4), 1946-1953.
- [15] Stavrakas, V., & Flamos, A. (2020). A modular high-resolution demand-side management model to quantify benefits of demand-flexibility in the residential sector. *Energy Conversion and Management*, 205, 112339.
- [16] Luo, X. J., Oyedele, L. O., Ajayi, A. O., & Akinade, O. O. (2020). Comparative study of machine learning-based multi-objective prediction framework for multiple building energy loads. *Sustainable Cities and Society*, 61, 102283.