

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 9, Issue 1, January 2020

## Survey on the Most Efficient Machine Learning Algorithm to Predict The Future Electricity Consumption And Production

Sohum R.Dhavale<sup>1</sup>, Sakshi D.Mane Deshmukh<sup>1</sup>, Khyati P.Shah<sup>1</sup>, Pradyot J.Itkurkar<sup>1</sup>

B. E Student, Dept. of Computer Engineering, NBN Sinhgad School of Engineering, Ambegaon (Bk), Pune, Maharashtra, India<sup>1</sup>

**ABSTRACT:** The forecast of electricity consumption is very important for policy making of the countries. Even after much research, the energy demand forecasting remains a problem. This paper presents a way to efficiently predict the future load on the system using different machine learning approaches. The collected datasets for electricity consumption pattern are from various domestic and commercial areas. Electricity consumption readings are collected by Axonet Smart Energy meter. To accomplish this, the survey introduces the most efficient algorithm among these algorithms K Means Clustering Algorithm, Naïve Bayes Classifier Algorithm, Support Vector Machine Algorithm, Decision Tree Algorithm, Random Forest Algorithm, Convolutional Neural Network (CNN), to forecast the consumption using the historical electricity consumption data.

**KEYWORDS:** Electricity Consumption- [n Prediction, Smart Energy Meter, Forecasting, K-Means, Support Vector Machine (SVM), Convolutional Neural Network (CNN), Naïve Bayes Classifier, Decision Tree, Random Forest.

### I. INTRODUCTION

In the recent years, the demand for electricity has increased in household as well as in industry with the use of different appliances. With the non-linear time series needs of electricity, a lot of electricity produced by the producer is wasted. Thus, the need arises to predict the future consumption of electricity. According to the fact that the use of electricity is fluctuating, the survey introduces prediction model of electricity by adopting historical electricity consumption data. The vital pillar of economic growth is energy. A good demand prediction is prerequisite for effective development of energy forecasting system. Therefore, accurate forecast of energy consumption is very crucial [1]. Increasing human population and large scale industrialization leads to change in energy rate of energy consumption.

#### I.(a) Load Forecasting

Forecasting of load is categorized in three categories short term, medium term and long term load forecasting [2]. There are two main approaches for load forecasting Physical Principle Based Forecasting and Statistical and ML based Forecasting. This paper works on second approach. In literature, different ML approaches are investigated. In [3], the paper investigate effectiveness of using Convolutional Neural Networks (CNN) for load forecasting. The paper compared CNN with Artificial Neural Network (ANN), Support Vector Machine (SVM) and other deep architectures. In [4], hybrid method using regression predictor and power peaks compensator was used to perform short term load forecasting. The method to predict thermal comfort sates of occupants was used by K-Means approach in [5]. In [6],

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 9, Issue 1, January 2020

Support Vector Machine (SVM) was used to perform long term load forecasting. This paper [7], uses Naïve Bayes classifier to predict overall energy generated by solar cells (photovoltaic cells).

## I.(b) Methodologies

This section provides a brief overview of different Machine Learning Approaches.

CNN's are one of the distinct type of Neural Network, which processes the data using grid topology [8]. Images can be viewed as 2D grid whereas 1D grid for energy consumption data. CNN's have mastered in Image classification for computer Vision tasks [9]. The specialized linear operation called 'Convolution' is used by CNN in at least one of the network layers.

Convolution is an operation performed on two functions on real valued arguments [9]. This operation is denoted by an asterisk.

$$s=(x*w)$$

where x represents input function and w represents waiting function. In CNN waiting function is termed as "kernel", and "feature map" denoted by s is the output of convolution operation [3].

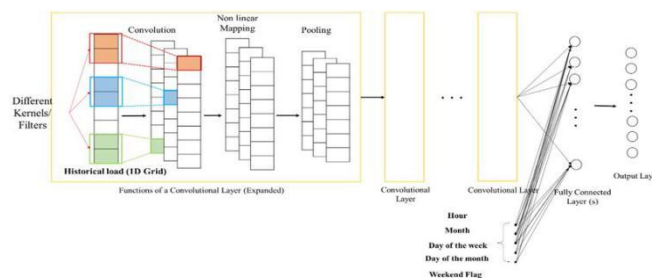


Fig -1: CNN architecture for energy load forecasting[3]

K-Means algorithm is used to partition the dataset into k predefined distinct non-overlapping clusters where each data point can lie under only one subgroup. It assigns data points to the specific cluster in a way that sum of the squared distance between the data points and the centroid of the cluster is at minimum distance [10].

K-Means algorithm work flow:

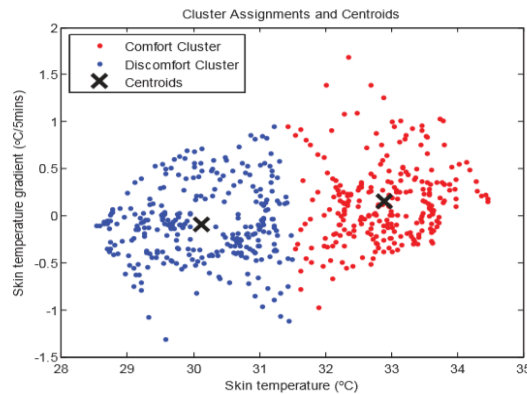
1. Define number of Clusters
2. Initialize centroids after dataset is shuffled, then randomly select k data points.
3. Until there is no change in the centroid repeat the above steps.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 9, Issue 1, January 2020



**Fig -2:** K-Means Clustering [5]

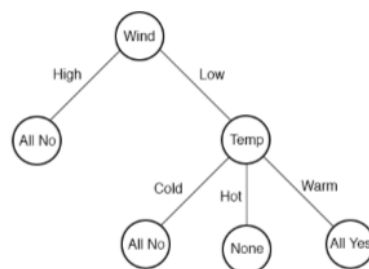
Naïve Bayes classifier is a simple Bayesian classifier which is widely used in the field of data mining. The Naïve Bayes classifier uses the assumption of conditional independence. That means, every attribute is conditionally not dependent on other attribute[13]. The Bayes’ Theorem is described below,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

**Fig 3-:** Naïve Bayes classifier

where,  $P(A/B)$  is the probability of event A in case B occurs,  $P(B/A)$  is the probability of event B in case A occurs,  $P(A)$  is the probability of A and  $P(B)$  is the probability of B [7].

Decision Trees are also termed as prediction tree, it refers to a tree like structure which is used to denote the sequence of decisions as well as consequences. In decision trees input variable are considered as categorical or continuous. It is possible to achieve the prediction by using the approach of generating tree which has test points and branches. Test points are nothing but a set of nodes which help in selecting the path to traverse the tree. Leaf nodes are the nodes which don’t have further branches, they return the class label value along with its probability score.



**Fig -4:** Sample Decision Tree

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 9, Issue 1, January 2020

## II. LITERATURE SURVEY

### II.(a) Analysis and predicting electricity energy consumption using data mining techniques.

[14] This paper shows that after the conducted analysis, regression model have more accuracy and less error rate than data mining techniques like Neural Nets and SVM models for predicting the electricity consumption. Predicting the electricity consumption is important, especially in developing countries for policy making. The predictive variables considered in dataset of consumption of electricity in Islamic Republic of Iran-Mazandaran province are population, moisture, temperature and price of electricity's unit. After building the desired models, the best model was selected based on correlation rates and relative error. The results of prediction showed that the electricity consumption will increase by 3.2% per annum on average. Electricity consumption rate was mostly affected by field population. More accurate predictions can be made using monthly or daily data. Concluding the most efficient algorithm amongst above mentioned three data mining techniques is the main motto of this research.

### II.(b) A Novel Application of Naïve Bayes Classifier in Photovoltaic Energy Prediction.

[7] This paper uses the Naïve Bayes classifier to predict the overall energy generated by photovoltaic cells which are installed in distributed region. The dataset includes the data values gathered at interval of one day. Based on rule based inferences, continuous-valued data is converted into categorical-valued data. Categorical valued data is represented as class labels like 'very high', 'high', 'medium', 'low', 'very low' and daily sunshine duration, global solar radiation, average temperature, photovoltaic energy generation attributes are considered as the continuous-valued data. Using input "daily photovoltaic energy generation" attribute and applying Naive Bayes classifier on the datasets, output attributes like daily "average temperature", "total global solar radiations " and "total sunshine duration" are predicted. The measures like 'Sensitivity' and 'Accuracy' of photovoltaic cells are improved by Naïve Bayes application. The Sensitivity and accuracy characteristics of the proposed Naïve Bayes algorithm are enhanced using different input parameters. This can be consider for future work.

### II.(c) Energy Consumption Load Forecasting Usinga Level-Based Random Forest Classifier.

[16] This research paper does the prediction using the pattern recognition. This scheme proposed to predict direct desired level in absence of regression using simpler classifier model. In forecasting the energy load the difficulty occurs that the forecasted load of energy of next hour might be low or high and the consumer is unaware of the same. The above mentioned method is the solution of the same problem. The Random forest algorithm which uses the various machine learning techniques is used in the method. In the dataset the data points are partitioned into the Ordinal bins with approximately equal number of data points. Data set contains the 3 energy levels which are preprocessed numerical values which is carried out by "machine learning random forest classifier". The method of random forest proposed in the paper is serves as the better alternative for the regression model and gives accurate results but it is time consuming system.

### II.(d)Accuracy Prediction for Distributed Decision Tree using Machine Learning approach.

[15]. In this paper, an improvised version of Distributed Decision tree Algorithm is proposed which will take less time for building the decision tree that too without compromising the accuracy of Decision Tree. This is implemented in Spark Environment's shared-nothing architecture. The datasets used were skin-Segmentation, wine quality, click on ad prediction and credit card fraud detection. Performance can be improved by making the tasks that are time consuming perform parallel and also make careful design such that it is scalable. Horizontal data parallelism is performed in the Spark Environment. Also, any algorithm, which is scalable can be implemented in Spark owing to its in-memory computation time and speed, map-reduce and Resilient Distributed Dataset. It can be concluded that the performance of Spark is better than Planet, an Hadoop Implementation but Speedup factor is one of its drawbacks.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 9, Issue 1, January 2020

## II.(e) Deep Neural Networks for Energy Load Forecasting.

[3] to solve the difficult problem of accurate load forecasting this paper proposed a methodology which is based on the deep neural networks to be specific this method uses convolutions on historical loads which is Convolution Neural Networks also known as CNN to predict used energy quantity on individual building level. The output includes other apposite information in fully connected layers together. This method is mainly used when the datasets to be fetched should be in grid format. The system relies on the main objective that the estimation of consumption of energy is for a time step or may be for multiple time steps in the upcoming future. In this research paper the work is the extension to their own work of deep neural network. To check the accuracy of CNN methodology it is also implemented with ANN methodology and SVM methodology for comparison after that different CNN architectures which has different convolution layers are tested. Author has taken the references of "Deep Belief Network". Whether input data points are the essential inputs for learning process of algorithm as it is much more related. As a result it is concluded that CNN methodology has more efficiency than SVM methodology and it performs equally as the ANN and deep neural network methodology.

## II.(f) Energy Efficiency Improvement with k-means Approach to Thermal Comfort for ACMV Systems of Smart Buildings.

[5] this paper uses k-means algorithm for the prediction of the thermal comfort states of the recipients. This approach is divided into 2 main tasks viz. Modeling and Optimization. Firstly, the modelings of skin and air temperature, total energy consumption as well as the skin temperature gradient are verified post their implementation. To solve the Optimization Problem, the undesired operating frequencies are filtered out using k-means approach and this approach also locates the optimal frequencies using the Augmented Firefly Algorithm (AFA). The results show that the energy efficiency, while maintaining the indoor temperature, is improved considerably, thus fulfilling the objectives of smart building. Limitations of this study is that the thermal comfort conditions set by the Air-Conditioning and Mechanical Ventilation (ACMV) Systems may not satisfy condition of every single room of the desired smart building.

## II.(g) Machine learning based Residential Energy Management System.

[17] In this paper to solve the problem of global energy crisis with climate change, the "Residential Energy Management System" is proposed. This system switches the energy to renewable local energy storage. System checks charged and discharged state and availability of the grids to reduce power consumption from grid. Artificial Neural Networks (ANN) and Support Vector Machine (SVM) machine learning algorithms are used for switchover. The energy load from the grid is effectively handled by switching the loads. In this paper the SVM model is implemented on the hardware platforms. When the above Residential Energy Management System is installed in every home community it will affect on large scale and show the significant reduction in the usage of loads from the grids

## III. EXISTING SYSTEM

Electric meters measure consumption of electricity. The traditional electricity meters consist of only rotors so they are less accurate and inefficient. They are not able to communicate with the environment i.e. the user or the energy provider. As compared to smart electricity meters, traditional meters provide less privacy and minimal security [12]. As the consumer may not know the future needs of their electricity consumption, the producers don't get any clear idea about the amount of electrical energy they should produce. This leads to huge amount of wastage of the electricity produced as well as the resources used in producing that energy. As examples of producers in the current system, we can consider the energy produced by windmills, coal power plant and hydroelectric power generators.

In the current system, the electricity providers' personnel has to manually collect the meter readings from sites wherever the traditional meters are installed. This leads to acute wastage of time, money and manpower. And with the

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 9, Issue 1, January 2020

increasing population, the need to predict the energy consumption is increasing drastically so that there is less wastage and efficient utilization of the produced electricity.

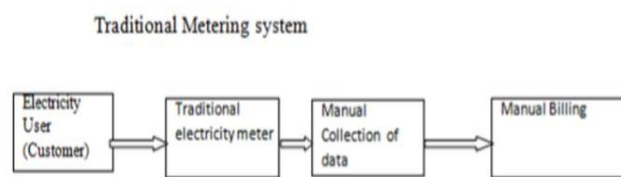


Fig -5: Traditional Electricity Metering System [12]

## IV. PROPOSED SYSTEM

Smart electric meters consists of chipset instead of the rotors which calculate the flow of current and then share the data with database which can be accessed by the customer as well as the energy provider. Manual collection of reading is not needed due to the smart meters. The Smart energy meters communicate with the environment with the help of WAN (wide area network), HAN (home area network). This network is used to allow meters to communicate with the users, as well as the energy provider. This system will be using this collected data to predict the future consumption of electricity. This will also give insight to the producer what will be the future requirement of energy.

### IV.(a) Working

The data which is being generated is collected by the Axonet Smart meters (smart energy meter). These smart meters are installed on-site. They are used in domestic sector as well as in industrial sectors. The meters are connected with the central system in spite of being geologically distributed. From there the raw data is sent to the Datonis API it is the platform through which the data is channelized and the data is managed using the SiteWhere Instance, this JSON parsed data is then stored in Amazon Web Server.

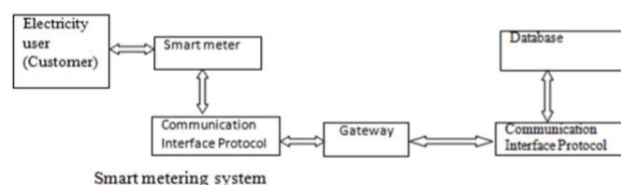


Fig -6: Smart Electricity Metering System [12]

The dataset from the AWS contains different readings collected by various smart energy meters. These readings specifies the amount of electric energy consumed by end user on particular day at particular amount of time. Preprocessing techniques such as data-cleaning, removing fuzzy/inconsistent data, data transformation, and data categorization are applied on this dataset for efficiently processing the data.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 9, Issue 1, January 2020

This dataset is partitioned into training dataset and testing dataset. Based on the training dataset provided to the algorithm (classifier) the prediction model is created. The model forecasts the future consumption of electricity by certain user using the historical data provided by the smart energy meters. It is beneficial for both electricity consumer and producer.

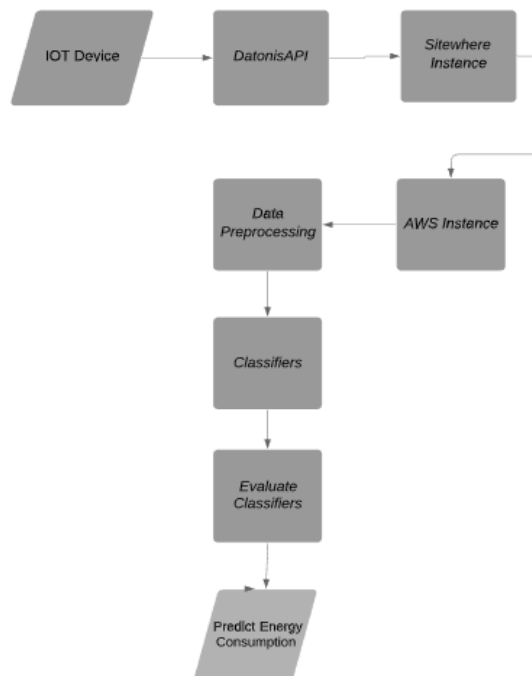


Fig -7: System Flow Diagram

Cross validation method is used for analyzing the performance of the model to estimate how accurately the model will work in real-time. The testing data is used to compute the confusion matrix. Confusion matrix is a tabular representation of the performance of the model. It shows the accuracy of the model in terms of Positive (P) and Negative (N) observation such that:

True-Positive (TP): Observation is positive and prediction is positive.

False-Positive (FP): Observation is negative but prediction is positive

True-Negative (TN): Observation is negative and prediction is also negative.

False-Negative (FN): Observation is positive but prediction is negative

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Fig -8: Classification Rate/Accuracy



# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 9, Issue 1, January 2020

## V. APPLICATION

This system provides the benefits to the consumer as well as producer, by analysis of historical data of electricity usage by consumer the producer can predict future consumption of electricity, this will help the producer to anticipate the future need which will reduce the wastage or shortage of electricity.

This helps the consumer to regularly monitor and analyze consumption patterns, which leads to altering the usage of electricity consumption. For commercial usage this systems helps the organization to efficiently plan the energy overheads for particular system or project.

## VI. CONCLUSION AND FUTURE SCOPE

This paper proposes the system to investigate the efficient Machine Learning approach for energy load forecasting.

In this system Support Vector Machine (SVM), K-Means clustering algorithm, Naive Bayes classifier, Random Forest algorithm, Decision Tree, Convolutional Neural Network (CNN) have been proposed for forecasting energy loads. The efficiency of the classifier is based on the matching percentage of the predicted and actual values. Forecasting energy consumption is important to overcome the wastage of energy while producing and also helps to monitor the commercial energy system.

This system will prove to be beneficial, as it will allow the producer to foresee the future forecast of energy consumption. Taking this into consideration producer can efficiently manage the energy consumption in such a way that there is neither shortage nor the wastage of energy.

## REFERENCES

- [1] C. J. Chang and C. D. Li and C. C. Chen and W. C. Chen, Forecasting short term electricity consumption using the adaptive grey-based approach—An Asian case, OMEGA journal, 40, 2012, 767-772.
- [2] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," Sustainable Energy, Grids and Networks, vol. 6, pp. 91–99, Jun. 2016
- [3] Kasun Amarasinghe, Daniel L. Marino, Milos Manic, "Deep Neural Networks for Energy Load Forecasting",Conference: 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)
- [4] X. Z. Gao, A. Kaarna, L. Lensu, S. Honkapuro, "A Hybrid Method for Short-Term Electricity Consumption Prediction", IEEE, IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society
- [5] Deqing Zhai, Tanaya Chaudhuri, Yeng Chai Soh, "Energy Efficiency Improvement with k-means Approach to Thermal Comfort for ACMV Systems of Smart Buildings", IEEE, 2017 Asian Conference on Energy, Power and Transportation Electrification (ACEPT)
- [6] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy Load Forecasting Using Empirical Mode Decomposition and Support Vector Regression," IEEE Transactions on Smart Grid, vol. 4, no. 1, pp. 549–556, Mar. 2013.
- [7] Ramazan Bayindir, Mehmet Yesilbudak, Medine Colak, Naci Genc, "A Novel Application of Naïve Bayes Classifier in Photovoltaic Energy Prediction", 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)
- [8] Y. Bengio, Y. Lecun, and Y. Lecun, Convolutional Networks for Images, Speech, and Time-Series. 1995.
- [9] I. J. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
- [10] Imad Dabbura, D(2018,September 17). K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks. towardsdatascience.com Retrieved from <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>
- [11] Samuel Bimenyimana, "Traditional Vs Smart Electricity Metering Systems: A Brief Overview", Journal of Marketing and Consumer Research, Vol.46, 2018
- [12] Le, T. N., Chin, W. L., Truong, D. K., & Nguyen, T. H. (2016). Advanced Metering Infrastructure Based on Smart Meters in Smart Grid. In Smart Metering Technology and Services-Inspirations for Energy Utilities. InTech.
- [13] S.S. Nikam, "A Comparative study of classification techniques in data mining algorithms", Oriental Journal of Computer Science & Technology, vol. 8, No. 1, pp. 13-19, 2015
- [14] Noorollah Karimtabar, Sadegh Pasban, Siavash Alipour, "Analysis and predicting electricity energy consumption using data mining techniques- A case study I.R. Iran - Mazandaran province ", 2015 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA 2015) March 11-12, 2015





ISSN(Online): 2319-8753

ISSN (Print): 2347-6710

# International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Visit: [www.ijirset.com](http://www.ijirset.com)

**Vol. 9, Issue 1, January 2020**

- [15] Siddalingeshwar Patil, Umakant Kulkarni, "Accuracy Prediction for Distributed Decision Tree using Machine Learning approach ", Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019), IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8
- [16] Yu-Tung Chen, Eduardo Piedad Jr., Cheng-Chien Kuo, "Energy Consumption Load Forecasting Using a Level-Based Random Forest Classifier", Creative Commons Attribution(CC BY) license (<http://creativecommons.org/licenses/by/4.0/>), 29 July 2019
- [17] Krishna Prakash N, Prasanna Vadana D, "Machine learning based Residential Energy Management System", 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)