Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July 2021: 3988 - 4000

Research Article

Energy Utilization and Prediction using Machine Learning for Improving EMS system: A Study Approach

[1] Dr.R.Muthukumar, [2,] L Vijay Anand, [3,] E.Sathish, [4,] E.Immanuvel Bright [5]R.Sathish.

[1]Associate Professor, Department of EEE, Erode Sengunthar Engineering College, Tamilnadu, India
[2, 3, 4] Assistant Professor, Department of EEE, Erode Sengunthar Engineering college, Tamilnadu, India
[5] Assistant Professor, Angel College of Engineering and Technology, Tamilnadu, India

ABSTRACT:
Construction Energy Management System (CEMS) has been an impressive topic these days on account of its importance in lessening energy wastage. In any case, the display of one of CEMS applications which is energy use conjecture has been lifeless due to issues, for instance, low assumption precision. Thusly, this assessment expects to determine the issues by cultivating a judicious model for energy use in cloud-based AI stage. Three frameworks which are Support Vector Machine, Artificial Neural Network, and k-Nearest Neighbor are proposed for the computation of the judicious model. Focusing in on certified application in India, two occupants from a business building are taken as a context oriented examination. The data assembled is examined and pre-arranged before it is used for model getting ready and testing. The introduction of all of the strategies is broke down subject to RMSE, NRMSE, and MAPE estimations. The experimentation shows that every inhabitant's energy use has different apportionment characteristics.

KEYWORDS: KNN-K-Nearest Neighbor, AI- Artificial Neural Network-EMS: Energy Management System RMSE- Root Mean Square Error; Mean Average Error (MAE)

1.0 INTRODUCTION

Lately, splendid design thought has been changed even more as frequently as conceivable as a drive to make a savvy space area by abusing the quick improvement of computational and correspondence designing [1]. This thought isn't just confined to India anyway various countries as well. Generally speaking populace understanding of clever structure thought turns on motorized cycle, which can thus control the construction's action using instrumentation measures and microcontrollers in two-way correspondence [2]. Other than mechanized control, a wise creating moreover includes a savvy system which gives energy usage guesses as an energy efficiency drive. This is a direct result of its advantage of yielding effective venture reserves and as a sensible philosophy for energy the board to restrict energy wastage [3].

A wise energy usage deciding is critical, especially for structures as designs' energy use is extending and

almost comes to 40% of fundamental energy use in made countries [4]. In India alone, energy usage has been augmentation little by little due to the improvement of people. The advancement of people lead to the growing of energy interest in this country and have been evaluated to show up at 116 million tons of oil reciprocals (mtoe) by this year. Energy gave in India is affected by the essential oil based commodity sources which included coal, combustible gas and fuel oil. Designs which including business, private and current in our country utilizes a total of 48% of the force that have been made [5]. The growing of energy uses towards structures starting with one day then onto the next make execution to this country in managing and lessening the energy usage whatever amount as could sensibly be relied upon to additionally foster energy usefulness.
This investigation is a proceeding with research from our past work where already measurable assessment and k-nearest neighbor (k-NN) method were proposed as the methodologies and SPSS was used as the stage [6]. In our past assessment, simply k-NN was proposed as the procedure to expect energy use. It is difficult to tell whether the strategy proposed was the great there is no connection had been made. Therefore, another two methodologies from AI are incorporated this assessment.

This investigation has utilized Machine Learning Studio, which is a web organization answer for the progression of assumption model. Starting from data assessment until execution evaluation, has been adequately used for the execution of energy demand deciding. A critical advantage of using over SPSS is it is not difficult to utilize and easy to use even the customer simply has essential data in appropriated registering and AI. One of the unmistakable features of was its ability to travel through a portrayal work measure. The work cycle that was coordinated inside the environment was controlled through a graphical improved on methodology. Other than that, parsing data for test was simply done by joining of modules. Besides, the stage similarly maintains script groups and computations written in external programming language, particularly R programming.

The headway of energy use insightful models that usage quantifiable examination and learning theory has a couple of colossal challenges. Attewell and Monaghan [7] portrayed that quantifiable conjecture is restricted especially by virtue of an immense dataset with a couple of features, as it requires a higher computational power for showing. Other than that, the verifiable assumption method itself is identically weak as it performs better by virtue of fixed time series and high comparability of usage level [8].

Newsham and Birt [9] also inferred that time series examination for power use guess execution was inadmissible due to factors in the picked credits. In addition, the standard headway of a judicious model is ordinarily established on the example of most noteworthy premium (kW) usage just, which is known as a period series procedure [10]. The model improvement would ignore other electrical limits, for instance, responsive power changes, which causes the model to be arranged unmistakably with chronicled data of most prominent interest regard. Out of the blue, the thought of various features of electrical power data would additionally foster the energy use estimate [11]. Subsequently, AI strategy is best when encouraging an insightful model of electrical use.

Changing from the quantifiable procedure to AI method itself doesn't handle all of the issues with energy use figure. Missing data that was accessible on a lot of data was outstanding to make the presentation of the perceptive model separate [12]. This missing data exists when in doubt due to the interconnectivity or

sensor issue which is the central multifaceted nature for sharp construction energy metering [13]. Moreover, the improvement of the AI model should utilize a cloud-based help to lessen the dependence of assumption on the gear points of interest [14]. Completely, the three fundamental spaces of energy use gauge discussed are AI figure method, treatment of missing data and work of cloud-based assumption showing stage which will be the reason of this investigation.

As the essential objective of this assessment is to encourage an energy usage perceptive model for canny business working by using a couple of AI methodologies in a cloud-based AI stage, this investigation zeros in extra on the precision of the methodology applied in expecting energy usage. Advances in AI looks at tremendously influence the field of adroit design energy the board as it is important to decrease energy use of various kinds of working from private constructions to mechanical constructions. In like manner, this examination is essential for the going with get-togethers, for instance, Ministry of Green Energy and Water and India Green Technology Corporations in their confirmation to separate energy usage level of the ebb and flow structure. It is similarly helpful to mechanical collecting association in expecting electrical stacking on their system for long arrive at projection, arranging of breaking point versus demand and for assembling plant improvement projection. To wrap things up, academician considering planning field to fathom the blend of data science and assessment in planning undertakings and high level training understudies to explore the organizations and likelihood of utilizing Machine Learning Studio for various exercises.

2.0 WRITING REVIEW

Computer based intelligence Prediction Methodology

Following the previously mentioned issue, this assessment keeps an eye on its hardships by driving assumption exhibiting through the evaluation of recorded power data.

This procedure utilizes a data driven technique as depicted by Corgnati et al. [15] whereby the data (regressor factors) and yield factors (response) are known. Considering this data, system limits will be surveyed and thus, a mathematical model could be created. A couple past assessments have researched the data driven AI approach. Fu et al. [16] proposed using one of ML computations which is Support Vector Machine (SVM) to predict the store at a design's system level (cooling, lighting, power, and others) taking into account environment gauges and hourly force load input. As a rule, SVM technique sorted out some way to expect the supreme force load with root mean square misstep (RMSE) of 15.2% and mean tendency bungle (MBE) of 7.7%. Revelations by Valgaev et al. [17] proposed a power demand figure using k-Nearest Neighbor (k-NN) model at a splendid construction as a component of the Smart City Demo Aspern (SCDA) project. The k-NN deciding technique was advanced toward using a lot of recorded discernments (consistently load twists) and their substitutions. The k-NN method is worthy at describing data yet confined in assessing future worth as it simply perceives practically identical cases in tremendous component space. In this manner, it ought to be enhanced with transient information recognizing verification whereby the assumption will be made for the accompanying 24 hours during workdays.

Five techniques for AI strategies were used for passing weight guaging by El Khantach et al.

[18] with a basic crumbling of the chronicled data done irregularly into time series of all day long, which finally settled 24-time series that tended to each earlier hour. The five AI methods used are multi-layer perceptron (MLP), support vector machine (SVM), extended reason work (RBF) regressor, REPTree, and Gaussian cycle. The experimentation was done ward on data got from the Moroccan electrical weight data.

The outcome showed that MLP philosophy came out as the most careful with MAPE level of 0.96 while SVM came straightaway and yet far from the result of MLP, it was still better contrasted with the rest. Yet the assumption for energy usage normally uses a request based AI system, estimate could in like manner be made ward on the backslide procedure as concentrated by González-Briones et al. [19]. The assessment constructed a perceptive model by taking apart the obvious instructive record using Linear Regression (LR), Support Vector Regression (SVR), Random Forest (RF), Decision Tree (DT) and k-Nearest Neighbor (k-NN). The limits of the investigation used one day-before power use (kWh) as an additional property. The results showed that LR and SVR models had the best execution with 85.7% precision.

The leading body of Missing Data

Techniques in dealing with missing data have been unfathomably thought already and methods have been inferred. There are two sorts of framework that are taking out the piece of the data which has missing worth and attribution method which relies upon close appraisal [20]. The essential strategy which block the missing snippet of data isn't feasible as this causes significant information to be taken out [21]. Without the data, a

attribution methodology is an optimal system. Newgard and Lewis [22] presented a couple of attribution techniques like Mean Value Imputation, Last Observation Carried Forward, Maximum Likelihood Estimate (MLE) and Multiple Imputation (MI). The mean worth attribution in a general sense substitutes the missing data with the mean worth of the dataset. In any case, this procedure isn't sensible for data which isn't absolutely subjective as it will introduce divergence in the data [23]. Another procedure presented was Last Observation Carried Forward, in which attribution is made for recorded data that was accumulated through [22].

The further evolved methodologies presented were Multiple Imputation (MI) and Maximum Likelihood Estimate (MLE) [22]. The Multiple Imputation procedure substitutes the missing data by bit by bit dislodging the missing data for each cycle made. This system utilizes quantifiable examination reliant upon saw data to manage the weakness that is introduced by the missing section. An outline of a standard MI system is Multiple Imputation Using Chained Equations (MICE) [24]. Most outrageous Likelihood Estimate conducts substitution through doubt made by from the start recognizing the limits and cutoff points reliant upon the dispersal of the data. The credit would then be made ward on the acknowledged limits. This procedure for credit was used by Probabilistic Principal Component Analysis (PPCA). Both advanced attribution methods have been broke down by Hegde et al. [20] in which attribution method was made on examined dataset containing 87 numeric-changed over hard and fast factors and 29 consistent

elements. The examination used RMSE estimations to evaluate attribution system execution. From the assessment, the PPCA methodology showed a much reassuring result appeared differently in relation to MICE, in which 65% of data factors were adequately credited by PPCA and simply 38% right attribution by MICE. This was also maintained by Schmitt, et al. [25] wherein the investigation dissected the display of six attribution procedures recalling PPCA and MICE for a certified dataset of various sizes. The result told that MICE sorted out the best way to perform well in a little dataset, yet in a huge dataset case, the MICE strategy performed deficiently.

3.0 Work of Cloud-based Prediction Modeling

There are diverse open cloud-based conjecture exhibiting stages which can maintain AI measure with additional abilities to inspect enormous data and streaming data. While picking the best AI gadget and stage, a couple of critical components ought to be considered, for instance, ascendable, pace, scope, practicability, versatility, and programming language [26]. Man-made intelligence stage and gadget are believed to be generally utilized with colossal data techniques for steady examination. Study of past papers on utilizing these ML stages with enormous data examination shows three top stages that are for the most part used for research which are Apache Spark's Machine Learning Library (Spark MLlib), TensorFlow, and Machine Learning Studio (ML). Apache Spark ML library gives a phase which backslide, gathering, and grouping [27].

Composing by Pérez-Chacón et al. [28] showed k-suggests estimation in Spark MLlib was used to see the electric energy use lead in a superstar series-based data. The results showed that the item sorted out some way to discover a day-based usage plan with low enlisting power for enormous data assessing to 3 years (2011, 2012 and 2013) for two constructions of a state subsidized school. Another sort of stage is TensorFlow, which is an open-source library, that is zeros in extra on significant learning and backing learning method. Under Apache 2.0 open-source approving, TensorFlow progression was then begun by Google Brain bunch. The item's name, TensorFlow, basically explains its construction whereby it executes a data stream graph, involves "tensors" (data bunch which will be taken care of) and "streams" (data development in the structure) [30]. On the workspace, it can utilize the two CPUs and GPUs resources, for instance, in the examination by Cai [31], where the investigation executed Convolutional Neural Networks (CNN) in TensorFlow. The computational stage which involves Intel Core i7 5820K and Nvidia GeForce GTX Titan X are both used to give a learning-based power and runtime showing.

The AI approach at the endeavor level has in like manner emerged unexpectedly due to the introduction of a flexible data framework for managing enormous data. These associations would regularly go for Machine Learning Studio () as it use cloud-based insightful assessment, as such requires less revenue in hardware for driving examination [14]. Such delineation of an association is British Petroleum (BP) whereby Artificial Intelligent is utilized to additionally foster their prosperity execution and work viability to the extent examining likely new energy by creating significant model inside lesser time [15]. Despite its convenience, maintains Python and R as its external programming . Likewise, ML stage can give AI organization from the start until the age of a judicious model and can continue with the resulting stage, whether or not circulating or passing on the model to a site or various stages [29].

Procedure

Considering the kept an eye on examinations concerning the essential locales, this assessment conducts energy usage assumption using a dataset as of late accumulated from a business building arranged in Klang Valley India, from June 2018 until December 2018. The business building is furnished with IoT meters that are related with the power delta connection at two critical occupants of the construction. Every tenant is apportioned into two locales including two IoT meters named occupant A1, A2, B1 and B2. Assembled data every second that planned into

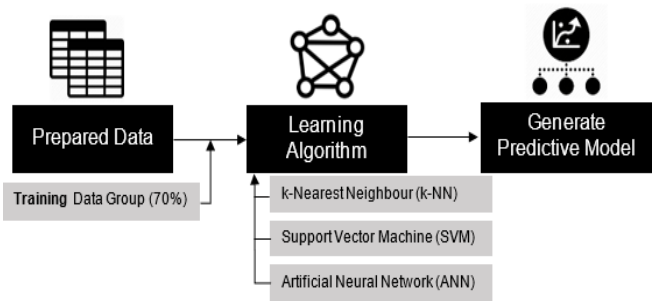

Figure 1 – Process of generating predictive model after data preparation

A.k-Nearest Neigbour (k-NN)

The essential insightful model to figure energy usage used the k-Nearest Neighbor (k-NN) method. This AI procedure is occasionally used due to its clear models and its assessing limit on confounding non-direct model [17]. It sorts out some way to give figure by choosing equivalent models between data centers in feature space [19]. For this investigation, the technique was used to predict most outrageous premium by using voltage, current and power factor as the features as the resultant expects a critical part in getting a strong match model to the data. Cost limit is quite far if the data point is misclassified or abuses most prominent edge.

B. Artificial Neural Network (ANN)

The third procedure in this assessment for energy use assumption was Artificial Neural Network (ANN). The advantages of using ANN like its ability to learn complex direct, makes it extensively used for conjectures and model affirmation . ANN model plan includes a course of action of interconnected neurons that have three essential layers; input layer, concealed layer, and yield layer. By differentiating the hidden yield and the best yield, change of the synaptic heap of every association that partners between the neurons was had until the impact is inconsequential (restricting Sum Squared Error (SSE)); this would offer regularization to the model [29]. The weight is the depiction of the need or meaning of the neuron input. For this assessment, a Multilayer Perceptron Model (MLP) sort of ANN structure with botch back inciting learning estimation was used for its association course of action structure. In the mysterious layer, a sensible non-direct trade work was used to calculate the information recognized by the data layer. The ANN model is as shown in condition.increase of the characteristics will yield the electrical power use

(kW). Considering the Euclidean distance work in condition 4, k-NN procedure was ready on and on up to the most outrageous tuning limit (k-regard) which reciprocals to 49. The resultant model from the readiness with the most diminished RMSE regard was picked for figure.

Euclidian distance work = J∑k (xi − yi) 2 (1)

C. Support Vector Machine (SVM)

In this assessment, the Support Vector Machine (SVM) was used with Radial Basis Function (RBF) as its segment work. This framework is regularly known as a biggest edge classifier and is utilized to deal with issues as for plan and backslide for an enormous dataset [36]. There are a couple of spot conclusions available for SVM procedure. In this assessment, Radial Basis Function (RBF) as shown in condition 5, was picked as a result of the wide and non-direct characteristics of the dataset.

RBF(x, xu) = exp[−y‖x − xu‖2] (2)

where y is a gamma limit to choose the spread scattering of the bit and ‖x − xu‖ is the Euclidean distance between the plan of core interests.

Condition 5 has a relating definition as tended to in condition 6 using sigma limit [17]

where m is the amount of data centers, n is the amount of concealed centers, f is the Sigmoid Transfer work,

{αj, j = 0,1, … n} is the vector of burdens from the mysterious layer to the yield layer and {þij, I = 0,1, … N; j = 0,1, … n } is the heap from the commitment to the mysterious centers.

For this investigation, the hyper parameter tuned was the amount of neurons per layer. This number of neurons implies the width of the association and its lethargic space [30]. Another rebuffing limit that was tuned and applied was weight decay. This limit is a rebuffing methodology to oblige the multifaceted design of the model and to confine the improvement of the model's weight limit [33].

Stage 4: Model Evaluation (Test)

Preceding contributing the data to the AI computation, the data was partitioned into two social occasions whereby 70% of the dataset was used for getting ready and the other 30% was divided trying data get-togethers. The readiness social occasions of data were used to set up every AI estimation and produce a perceptive model that could yield regard that matches with the recorded most outrageous interest data while the rest of the data was held down to be used to test the pre-arranged farsighted model. The cycle is as displayed in Figure 2.

With , data dividing getting ready and testing would not be an issue and uneven as it has basic assistance for data division. The allotting cycle was immediate in which decision was made randomly. This cycle hindered over fitting, which could cause There are 2 tuning limits for SVM-RBF which are bit limit

sigma (a) and cost limit (C), either underestimation or overestimation of the best advantage regard. its resolute quality and likeness to expect energy usage

4.1 Normality Testing of Dataset

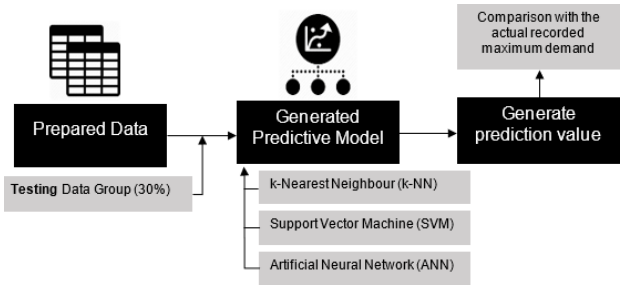

Figure 2 – Testing of the trained predictive model

During model setting up, a couple of models were made with different tuning limits for each method, in which k-regard was adjusted to k-NN tuning; sigma and C limit were changed for SVM-RBF tuning; and weight decay and mystery unit size were adjusted to ANN-MLP tuning. After the kept changing finished to its individual most prominent limits, each model was evaluated reliant upon Root Mean Square Error (RMSE), R-Squared (R2) and Mean Average Error (MAE). The formula is as shown in conditions 8, 9 and 10, independently, given that At is the certifiable recorded potential gains of most prominent interest data, and Ft is the expected characteristics. Yet 3 evaluations were made, just RMSE result was perceived as the best model for each method.

For the conventionality preliminary of the energy demand data, the skewness and kurtosis values were resolved using the complete data for every inhabitant, starting from June 2018 until December 2018. This evaluation was intended to analyze whether the condition of data impacts the introduction of the made insightful model. The made data was assembled in Table 1 and Table 2, for skewness and kurtosis, separately. Also, Figure 3 shows the sort of the dataset for a graphical evaluation of conventionality. The skewness and kurtosis values were figured for all credits, including the interest.

After the farsighted model using every AI estimation was made and assumption demand data was created, they were then surveyed to choose their show and accuracy. Three systems for evaluation were used which were Root Mean Square Error (RMSE), Normalized RMSE and Mean Absolute Percentage Error (MAPE). The formula for RMSE is as shown in condition 8 and MAPE in condition

Relationship of execution of the methodologies to different inhabitants was made by using a normalized RMSE or in any case called Coefficient of Variation RMSE (CV RMSE). This estimation dispenses with

the scale ward of RMSE [42].

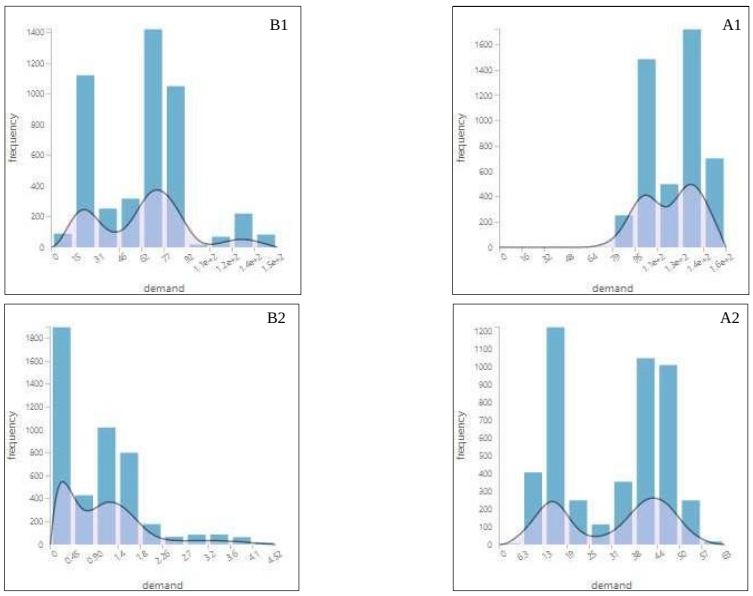

Figure 3 - Probability thickness for occupant B1, B2, A1 and A2

4.0 RESULTS AND DISCUSSION

The results of the experimentation were discussed in fragments subject to the method for the figures' framework. The pre-dealing with methodology and attribution of missing data using modules given by both Machine Learning Studio and Caret Package were discussed the extent that framework, effects, and importance. The revelations with respect to energy use assumption were explored for every tenant and execution assessment was obliged the figure delayed consequence of Artificial Neural Network (ANN), Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN). Extensively, utilization of ML

Table 1 – Skewness for each inhabitant utilizing total information

Occupant	Power Factor	Current	Voltage	Demand
A1	-0.56618	-0.1754	-.2488	-.34978
A2	-1.67496	-.10242	-.03279	-.28883
B1	-1.98888	-.64522	-.25056	0.17222
B2	1.84686	0.3358	- .89777	1.41075

Considering Table 1 and Figure 3, Tenant A1, A2, and B1 dataset was around uniformity and skewed

with bimodal shape thickness. Taking everything into account, for Tenant A1, the thickness was skewed left, as the long tail featured the left however for Tenant B1, the thickness was skewed right. Inhabitant A2 shows imbalance normal whereby the tails between each end were around changed. Not equivalent to Tenant B2, the scattering was significantly

The thickness plot of Tenant B2 scattering shows that the thickness was also bimodal with right-skewed.

To the extent kurtosis, from Table 2, Tenant A1, A2 and B1 had excess kurtosis under 0. This suggests that the scatterings were platykurtic. This brand name was also seen reliant upon Figure 3 (A1, A2, B1) whereby the probability thickness plot has a more broad tail, and the apex place was more broad. Contrary to Tenant B2, the excess kurtosis was higher than 0 at 1.824909. This shows that Tenant B2 scattering was leptokurtic and had a higher change. From this commonness testing, Tenant A1 and A2 had a generally customary transport. In any case, Tenant A1 has a mean worth not actually the center. Occupant B1 had moreover an around regular transport yet had to some degree higher skewness and kurtosis appeared differently in relation to Tenant in office A. Occupant B2 dataset was significantly skewed and had a higher change in relationship with various inhabitants.

5.0 CONCLUSION

The course of this investigation has focused in on encouraging an energy usage farsighted model for two business divisions that have accepted the sharp design natural framework. The energy demand data accumulated from 90 days has been researched and pre-ready for planning and testing of the perceptive models. Utilizing the Machine Learning studio authentic examination of the data assembled was made to choose the normality of the dataset. From this examination, skewness and kurtosis values were acquired and set up that all assembled data was different in transport ascribes. Continuing with the cooperation of insightful model development, the data accumulated was pre-taken care of through the attribution of missing data using PPCA strategy and standardization change. The pre-dealing with was adequately executed inside environment, in which the resultant arranged data had a mean regarded at 0 and a standard deviation of

An experimentation cycle was similarly made to choose the capacity of the missing data attribution method. An investigated data with missing worth was made to survey the PPCA procedure, where it conveyed a promising result with low unrefined inclination and consideration speed of more than 90% of the genuine worth.

Focusing in on the objective of this assessment, three coordinated AI assumption strategies specifically k-Nearest Neighbor, Support Vector M9achine with Organization with Multilayer Perceptron model, were picked as the computation for the judicious model. These procedures were successfully examined similar to their resultant plan and assumption execution. The aftereffect of the model planning and testing shows that each strategy performed contrastingly for every tenant. SVM method shows the most promising result, whereby it sorted out some way to be the best system for 2 inhabitants which were Tenant A1 and Tenant A2, with RMSE regarded at 4.85 and 3.65 in the group independently. Besides, SVM result furthermore shows a lower mean preeminent misstep for Tenant B1 and Tenant B2 at 12.09 and 43.97,

separately, yet k-NN had lower RMSE result for these two occupants. SVM expected interest moreover would be savvy to precision when typical usage was resolved from the interest, where it achieved a lower MAPE than the rest of the methods for all occupants. Taking everything into account, this SVM ideal result goes with an expense, whereby the model made with the estimation took 12 to 18 hours to get ready. Alternately with k-NN methodology that performed to some degree more unfortunate than SVM, it just required most outrageous 40 seconds to get ready. For this investigation, ML was used for all cycles and R composing PC programs was used comprehensively for data normalizing, model getting ready and testing ultimately execution evaluation.

Cloud-based farsighted model improvement partakes in its advantage as it doesn't depend upon the introduction of the gear it is running on. Furthermore, it could sort out some way to hold back from bombarding due to sudden hardware conclusion. In any case, particularly doesn't have the right estimation needed for this assessment; consequently, R composing PC programs was used for the model planning and testing and execution appraisal. Completely, this assessment was triumphantly coordinated and the objectives were cultivated.

REFERENCES

1. B. Qolomany, A. Al-Fuqaha, A. Gupta, D. Benhaddou, S. Alwajidi, J. Qadir and A. C. Fong, Leveraging Machine Learning and Big Data for Smart Buildings: A Comprehensive Survey, IEEE Access 7 (2019) 90316-90356.
2. D. Xu, Z. Li, S. Yang, Z. Lu, H. Zhang, W. Chen, Network for Predicting Electric-Power Load, *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)* 1-6.
3. U. Berardi, Building Energy Consumption in US, EU, and BRIC Countries, Procedia Engineering 118 (2015) 128-136.
4. J. S. Hassan, R. M. Zin, M. Z. A. Majid, S. Balubaid, and M. R. Hainin, Building energy consumption in India: An overview, Jurnal Teknologi 707:7 (2014) 2180-3722.
5. N.L. Mazlan, N.A. Ramli, J.L. Awalin, M.B. Ismail, Kassim, and A. Menon, A Smart Building Energy Management using Internet of Things (IoT) and Machine Learning, Test Engineering and Management 83 (2020): 8083-8090
6. P. Attewel and D. Monaghan, Data Mining for The Social Sciences: An Introduction, Oakland: University of California Press, 2015.
7. S. A. Abdul Karim and S. Alwi, Electricity load forecasting in UTP using moving averages and exponential smoothing techniques, Applied Mathematical Sciences 7 (2013) 4003-4014.
8. G. R. Newsham and B. J. Birt, Building-level occupancy data to improve ARIMA-based electricity use forecasts, in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building (BuildSys '10)*, New York, NY, USA, 2010.
9. Z. Xiangyu, M. Qingqing, L. Tao, Z. Lei and Z. Jianquan, Linear Regression Electricity Prediction Method Based on Clustering of Electric Characteristics, in *ICBDT2019: Proceedings of the 2nd International Conference on Big Data Technologies*, Jinan, 2019.
10. R. Wei, J. Wang, Q. Gan, X. Dang, and H. Wang, Predicting Electricity Usage Based on Deep Neural Network, in *2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, Tianjin, China,

2019.

11. M. W. Ahmad, M. Mourshed, D. Mundow, M. Sisinni, and Y. Rezgui, Building Energy Metering and Environmental Monitoring - A State-of-the-art Review and Directions for Future Research 120 (2016) 85-102.
12. H. Nugroho and K. Surendro, Missing Data Problem in Predictive Analytics, in *8th International Conference on Software and Computer Applications (ICSCA 2019)*, Penang, India, 2019.
13. M. Mateev, Management of Smart Energy Systems with IoT-based Predictive Analytics for More Efficient Energy Consumption, in *ATINER's Conference Paper Proceedings Series: COM2019-0163* , Athens, 2019.
14. S. Corgnati, T. Bednar, Y. Jang, H. Yoshino, F. M.,
15. S., D. N., S. M., C. Ghiaus, C. A., T. N. and &. F. V., Total energy use in buildings. Analysis and evaluation methods, Final report Annex 53. Statistical Analysis and Prediction Methods, 2013.
16. Vector Machine to Predict Next Day Electricity Load of Public Buildings with Sub-metering Devices, Procedia Engineering 121 (2015) 1016-1022.
17. O. Valgaev, F. Kupzog, and H. Schmeck, Low- Voltage Power Demand Forecasting Using K-Nearest Neighbors Approach, IEEE Innovative Smart Grid technologies - Asia (ISGT - Asia) (2016)1019-1024.
18. El Khantach, M. Hamlich, and N. E. Belbounaguia, Short-term load forecasting using machine learning and periodicity decomposition, AIMS Energy 7:3 (2019) 382-394.
19. González-Briones, G. Hernández, J. M. Corchado, S. Omatu, and M. S. Mohamad, Machine Learning Models for Electricity Consumption Forecasting: A Review, IEEE Virtual-Ledgers-Tecnologías DLT/Blockchain y Cripto-IOT *(*2019).
20. H. Hegde, N. Shimpi, A. Panny, I. Glurich, P. Christie and A. Acharya, MICE vs PPCA: Missing data imputation in healthcare, Informatics in Medicine Unlocked (2019).
21. A. Manly and R. S. Wells, Reporting the use of Multiple Imputation for Missing Data in Higher Education Research, Research in Higher Education 56:4 (2015) 397-409.
22. D. Newgard and R. J. Lewis, Missing Data: How to Best Account for What is Not Known, Journal of the American Medical Association 314:9 (2015) 940-941.
23. H. Kang, The prevention and handling of the missing data, Korean journal of Anesthesiology 64:5 (2013) 402-406.
24. M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, Multiple imputation by Chained Equation: What is It and How Does It Work? International Journal of Methods in Psychiatric Research 49 (2011) 40-49.
25. P. Schmitt, J. Mandel, and M. Guedj, A Comparison of Six Method s for Missing Data Imputation, J. Biomet Biostat 6:1 (2015).
26. T. Landset, T. M. Khoshgoftaar, A. N. Richter, and
27. T. Hasanin, A Survey of Open Source Tools for Machine Learning with Big Data in the Hadoop Ecosystem, Journal of Big Data 2 (2015) 24.
28. J. Quddus, Machine Learning with Apache Spark Quick Start Guide, Packt Publishing, 2018.
29. Dr.A.Suresh, Saiyed Falayaz Waris "An Efficient Patch Based Local Principal Component Analysis Technique In Image Denoising Process With Soft Computing Approach" International Journal Of Scientific & Technology Research Volume 9, Issue 01, January 2020 ISSN 2277-8616
30. R. Pérez-Chacón, R. L. Talavera-Llames, F. Martínez-Álvarez, and A. Troncoso, Finding Electric

Energy Consumption Patterns in Big Time Series Data, in *Distributed Computing and Artificial Intelligence, 13th International Conference*, 2016.

31. B. Ramsundar and R. B. Zadeh, TensorFlow for Deep Learning, 1 ed., Sebastopol, California: O'Reilly Media, 2018.
32. Abadi, Martín et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *ArXiv* abs/1603.04467 (2016)
33. Dr.S. Vinoth Kumar,Dr. H.Shaheen,Dr. T. Sreenivasulu,An Improved Scheme for Organizing E-Commerce-Based Websites Using Semantic Web Mining,Intelligent Computing And Applications,pp 115-123,Springer, Singapore,2020