IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Energy Consumption Prediction Model For Residential Buildings Using Deep Learning And Machine Learning.

Kaveri Nandre, Department Of Computer Science , Amrutvahini College Of Engineering, Sangamner
Aachal Patil, Department Of Computer Science ,Amrutvahini College Of Engineering, Sangamner
Oshin Sharma, Department Of Computer Science , Amrutvahini College Of Engineering, Sangamner
Sahil ,Department Of Computer Science , Amrutvahini College Of Engineering, Sangamner
Guided By: Prof. S. A. Thanekar, Amrutvahini College Of Engineering, Sangamner

ABSTRACT

The high proportion of energy consumption in building leads to the manifestation of many en vironmental problems, which negatively affect the existence of human beings.Predicting ener gy consumption in buildings is basically stated as a method of saving energy and improving d ecision making for reducing the loss of energy consumption. The construction of energy efficient buildings will also contribute reducing the overall energy consumption in new buildings. Machine learning methods are considered to be the most appropriate method to deliver the expected result in prediction tasks.Therefore ML has been used in the field of the energy consumption of operational buildings in studies.

KEYWORDS

Building energy consumtion,Energy prediction,Machine learning,Energy efficiency

1.Introduction

Energy-efficient and feasible buildings have ended up basic buildings for protecting the environment, as building wastefulness is the driving cause of worldwide vitality utilization and nursery gas emissions. The tall vitality utilization of buildings leads to major natural issues such as climate change,air contamination and warm contamination, which have a genuine affect on human survival. The vitality request of buildings has expanded essentially over the past decades due to populace development and quick urbanization. Predicting the vitality utilization of buildings is fundamental for vitality investment funds and way better choice making to decrease vitality utilization. Coordinate modeling, moreover known as physics-based modeling strategies, frequently requires point by point data almost the building and its environment, such as HVAC (warming, ventilation and discuss conditioning) systems,floor thickness cover, warm execution, inside inhabitance loads, sun powered data, etc. Based on reenactment devices there are primarily DOE-2, Vitality Furthermore and TRNSYS with respect to this method. The demonstrate require parameters as well numerous, as a rule blocked off. In this manner, these apparatuses are considered wasteful due to the deficiently sum of data required and their time utilization. The show employments machine learning calculations for building vitality estimation and has been proposed in a few thinks about since it does not require a huge sum of point by point building input. This approach is prepared on a expansive dataset of point by point hourly or intra-hourly readings extricated from building administration frameworks and shrewd meters. The precision of these models in foreseeing vitality utilization of buildings depends on three factors:the demonstrate chosen, the amount and quality of the information.

2.Literature Review And Objective

2.1 Objective

To extend the effectiveness of buildings by decreasing the generally vitality utilization in buildings and consequently to contribute in sparing environment from worldwide warming.

2.2 Literature Review

Sr.No.	Paper Title	Year of publication	Algorithm used
1	Evolutionary Deep Learnibg Based Energy Consumption Prediction Of Buildings	2019	1)Long short term memory 2)Genetic algorithm
2	Machine Learning For Energy Consumption Prediction and Scheduling in smart buildings	2020	1)ANN 2)Genetic algorithm
3	Building model based onRough Set Theory and Deep Learning Algorithms.	2021	1)Rough set theory 2)SVM 3)ANN
4	Building Energy Prediction using deep learning and machine technique	2022	1)ANN 2)SVM 3)Linear regression 4)Decision tree

3.Methodologies of problem solving and efficiency issues

3.1 Random forest

Random Forest is an ensemble technique t hat provides a variety of beneficial features. These features include ; (i) It relie s on theIntegrated Learning Theory, which enables it to learn both simple and all questions. (ii) Compared to other ML algor ithms (such asartificial neural network, sup port vector machine, etc.), it does not requi re much hyperparameter tuning to get good performance from(iii) Standardsettings ge nerally provide excellent performance. The refore, random forest (RF) methods have r eceived increasing attention in the field of building energy consumption .

3.2 ANN

ANN called as the manufactured neural systems as a rule essentially neural networks(NNs) are computing frameworks propelled by the organic neural systems that constitute creature brainsAn ANN is based on a collection of associated units or hubs called counterfeit neurons, which freely demonstrate the neurons in a natural brain. Each association, just like the neural connection in a natural brain, can transmit a flag to other neurons. An manufactured neuron gets signals at that point forms them and can flag neurons associated to it. The "signal" at a association could be a genuine number, and the yield of each neuron is computed by a few non-linear work of the entirety of its inputs. Neurons and edges ordinarily have weight that alters as learning continues. The weight increments or diminishes the quality of the flag at a association. Neurons may have a limit such that a flag is sent as it were in the event that the total flag crosses that threshold. Different layers may perform distinctive changes on their inputs. Signals travel from the primary layer (the input layer), to the final layer (the yield layer), conceivably after navigating the layers numerous times.

3.3 KNN

The K-Nearest Neighbors (KNN) algorithm is a supervised ML algorithm that can be used for classification and estimation problems. However, it is frequently used in classification estimation problems in businesses. The K-Nearest Neighbors (KNN) algorithm uses "feature similarity" to estimate the value of the new data point, meaning that in training new data will be assigned a value based on how it matches the points. It is useful for non-linear data because the algorithm does not make any assumptions about the data.

3.4 DNN

The deep neural network model originally emerged from neurobiology. At a high level, the biological system receives many signals from its connections with its dendrites and releases a stream of influence from its axons. By splitting the access model, the complexity of multiple access can be reduced. Inspired by this theory, the neural network design consists of units that connect various inputs and form an output.

4 Model development

4.1 Model Evaluation

The execution of each demonstrate is assessed utilizing the taking after execution measures: R-Squared (R2), Cruel Outright Blunder (MAE), Root Cruel Squared Mistake (RMSE), Cruel Squared Mistake (MSE). Among all the recorded assessment strategies, the foremost frequently utilized for vitality utilization expectation are the MSE and RMSE.

4.2 Mean Absolute Error (MAE) may be a strategy of calculating the distinction between the anticipated values and the genuine values at each point in a diffuse plot. The closer the score is to zero, the way better the execution whereas the higher the score, the more regrettable the execution. It is computed as the normal of the supreme mistakes between anticipated and genuine exertion.

4.3 Mean Squared Error (MSE) is the degree of squared variety between the assessed values and the genuine values. MSE is an appraisal of the quality of a indicator. Models with mistake values closer to zero are consider the way better estimation show. It is additionally known as Cruel Squared Deviation (MSD).

4.4 Root Mean Squared Error (RMSE) is additionally a metric utilized to calculate the contrasts between evaluated esteem and the real seen esteem of the demonstrate. It is accomplished through the square root of the Cruel Square Mistake (MSE).

4.5 R-Squared (R2) could be a factual degree that determines the extent of the distinction within the target variable that can be defended by the autonomous factors. It shows the degree to which the data fits the show. R2 can create a negative result, be that as it may, the finest result of R2 is 1.0. It is additionally known as the Coefficient of assurance.

4.6 Data Collection

There are two sorts of information utilized for show advancement to be specific building metadata and meteorological information. The building related dataset was collected from the Service of Lodging Communities and Nearby Government (MHCLG) store. This information contains the metadata and vitality utilization information of 5000 diverse sorts of private buildings within the UK

4.7 Building Metadata

The building metadata for 5000 private buildings found inside ten zone postcodes within the UK were collected. The information comprise of 500 private buildings from each of the ten diverse postcode regions (Blackburn, Blackpool, Darlington, Yorkshire, Halton, Hartlepool, Frame, Middlesbrough, Redcar Cleveland and Warrington) for a clear and less equivocal comparison. The building metadata comprised of as it were parameters that can be recognized and adjusted amid the plan organize. This included floor level, roof portrayal, dividers depiction and Number of Tenable Rooms among others as appeared in Table 2 underneath. These were considered as the autonomous factors whereas the vitality information was considered as the subordinate variable. The dataset incorporates the yearly vitality utilization of each building for the year 2020. The input factors such as divider and windows sort or portrayal are considered vital factors as they have impact on the vitality utilization of buildings .. For occurrence, the suitable choice of divider or window sort can significantly decrease vitality cosumption .

4.8 Meteorogical Data

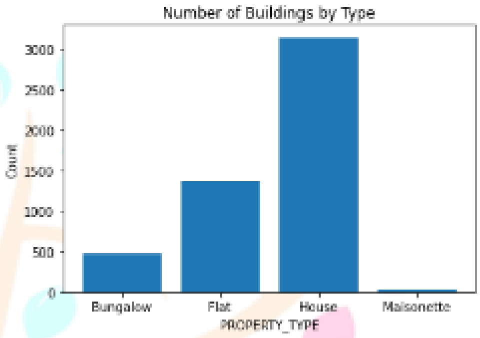
The meteorological dataset was collected from Meteostat store, and it contains climate highlights such as temperature, wind speed and weight as appeared in Table 2 underneath. Concurring to Ref. [72]; one of the major factors for building vitality expectation is meteorological information [72]. This information was collected for ten range postcodes of the private buildings and the granularity of meteorological information collected was every day normal from January 1, 2020 till December 31, 2020. Taking into thought the appropriateness of this show past UK, the meteorological information was found the middle value of month to month instead of yearly in correspondence to the vitality utilization information (such that the show wins in both areas with tall and direct climate conditions).

4.9 Data Pre-processing

In machine learning, the introductory prepare incorporates information pre-processing for the planning of information, in spite of the fact that, it is frequently time devouring and computationally costly . This prepare is vital to identify the presence of invalid or conflicting information that can cause blunder amid examination.

5.Data Merging

The utilization of numerous datasets requires information consolidating. Hence, the building dataset and meteorological dataset were combined utilizing the common variable (postcode) to coordinate each building information to its particular meteorological information. The datasets were consolidated utilizing the panda bundle of the python programming dialect. The information consolidated come about in a add up to of 60,000 information focuses.



5.1 Data Cleaning:

The method of information cleaning connected includes the expulsion of exceptions and treatment of lost information. The meteorological information contained few lost values which was settled by applying the cruel esteem and dropping the invalid esteem column.

5.2 Data conversion:

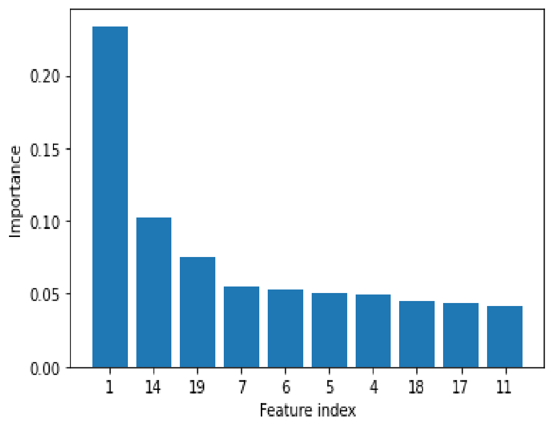
The building crude information comprised of a few categorical information in factors such as, divider vitality proficiency, windows vitality productivity among others .The factors were distributed values [e.g., exceptionally great = most noteworthy esteem (5) whereas exceptionally destitute = least esteem (0)] to supply a appropriate information for the ML calculation.

5.3 Data Normalization:

Information normalization could be a exceptionally common method of information pre-processing that dispenses with the impact of measurements as a few highlights regularly have disconnected measurements . Normalization scales person tests into a unit standard in arrange to dodge issues amid show development.The building dataset and the meteorological dataset were normalized utilizing sklearn python package normalizer. Sklearn python bundle may be a machine learning library for the python programming dialect that contains different algorithm's capacities counting pre-processing procedures such as normalization and standardization among others . Normalization utilizes the equation underneath to scales down the dataset such that the normalized values drop inside the extend of and 1.

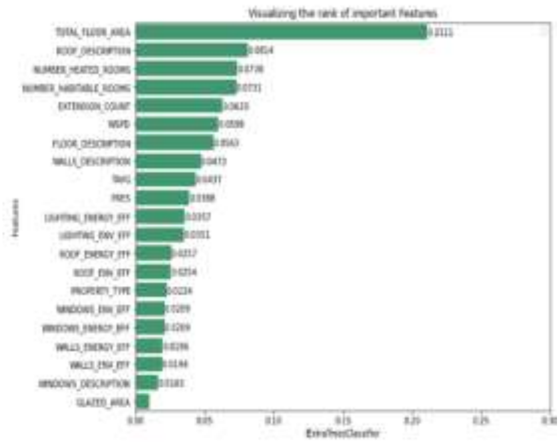
5.4 Feature Selection:

Feature selection (FS) is related to the exactness and complexity of prescient models. FS is an imperative prepare in executing MLmodel advancement since not all highlights have an affect. It is additionally vital to apply highlight choice for ideal execution of the .Arbitrary Woodland recursively analyzes the impact of including or evacuating highlights on the show and returns the relative reliance values for each variable.



5.5 Model Evaluation:

The execution of each demonstrate is assessed utilizing the taking after execution measures: R-Squared (R2), Cruel Supreme Mistake (MAE), Root Cruel Squared Mistake (RMSE), Cruel Squared Mistake (MSE). Among all the recorded assessment strategies, the foremost frequently utilized for vitality utilization forecast are the MSE andRMSE[1,21,23].

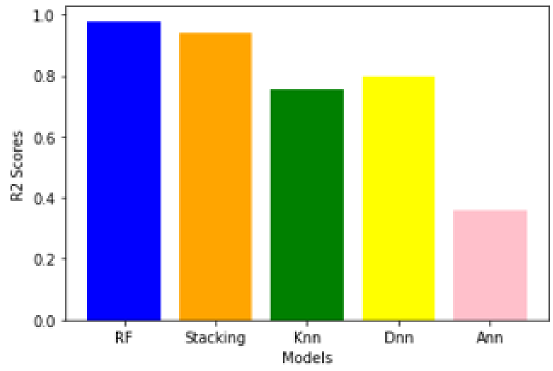


6.Result and discussion:

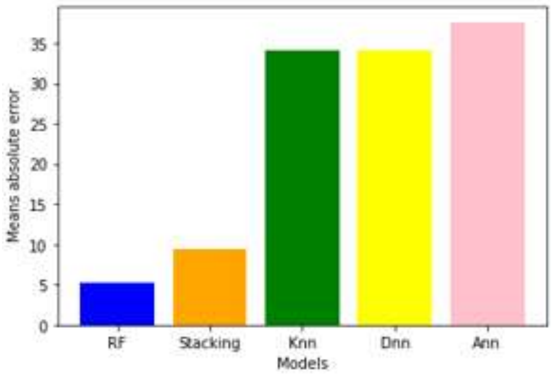
In this study, the examination of the result has appeared major discoveries between the chosen models. The premise of deciding the finest prescient demonstrate expressed in area 3.5 stipulates that show holding values closer to zero for MAE, MSE and RMSE are the great prescient show whereas values closer to one for R2 delivered the most excellent comes about. In this investigate, Irregular woodland developed the foremost productive show proficient Show for anticipating yearly vitality utilization. ANN based demonstrate, which is recognized for creating great comes about in expansive dataset has won. DNN, ANN and RF based models are specifically comparable. In spite of the fact that, it takes a longer to prepare the show.

Performance result for each model

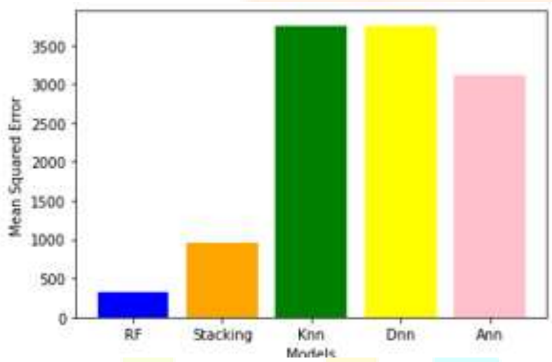
Model	Training time	R squared	MAE	RMSE	MSE
RF	7s	0.9789	5.3020	17.93	321.65
STACKING	4m	0.9374	9.3513	30.94	400.56
DNN	4m	0.7961	34.0525	61.275	3754.6
KNN	1s	0.7569	34.0525	61.27	3754.6
ANN	24s	0.357	37.5839	55.862	3120.6



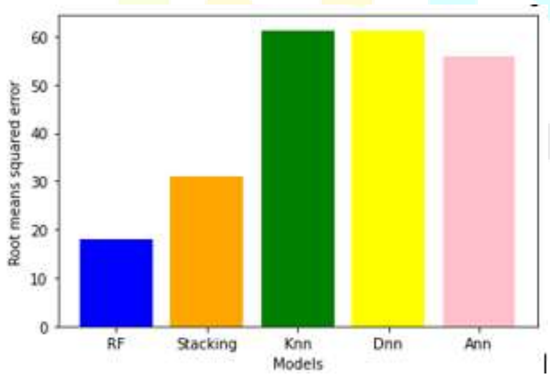
R-Squared Comparison



MAE Comparison



MSE Comparison



RMSE Comparison

6.Conclusion

We compare the expectation execution of nine machine learning strategies such as DNN, ANN, SVR, DT, GB, LR, Stacking, RF and KNN in yearly vitality utilization estimating. The positive comes about delivered within the consider give a demonstrate that permits creators to assess vitality utilization amid the starting plan stage. In general, DNNs created superior comes about for vitality estimation than other models. In any case, there are other estimation models such as ANN, GB and SVM that are viable in this consider. In terms of computational productivity, the DT performs best with a learning time of 1.2 s. A affectability analysis was at that point conducted to realize the third objective. Feature selection was performed utilizing the foremost imperative bunches within the writing, and nine machine learning models were made and analyzed. The comes about appear that the execution of the show isn't delicate to certain building sorts. In this manner, the building bunch does not have a critical affect on the chosen highlights and execution. In expansion, nine machine learning models were created utilizing two diverse datasets to meet the fourth objective.The comes about appear that the information measure has an effect on the execution of the show. The viability of this demonstrate empowers creators to utilize it in early plan to create educated choices, control and optimize the plan. Future investigate ought to center on utilizing DNN and other profound learning models for datasets bigger than 40,000 in trusts of way better execution. The appropriateness of other clustering calculations to anticipate building vitality utilization has moreover been examined.

7.Refetrences

[1] A.-D. Pham, et al., Predicting energy consumption in multiple buildings using

machine learning for improving energy efficiency and sustainability, J. Clean.

Prod. 260 (2020) 121082, https://doi.org/10.1016/j.jclepro.2020.121082

[2] Soheil Fathi, et al., Machine learning applications in urban building energy

performance forecasting: a systematic review, Renew. Sustain. Energy Rev. 133

(2020) 110287, https://doi.org/10.1016/j.rser.2020.110287.

[3] G. Serale, M. Fiorentini, M. Noussan, 11 - development of algorithms for building

energy efficiency, in: F. Pacheco-Torgal, et al. (Eds.), Start-Up Creation, second ed.,

Woodhead Publishing (Woodhead Publishing Series in Civil and Structural

Engineering), 2020, pp. 267–290, https://doi.org/10.1016/B978-0-12-819946-

6.00011-4.

[4] L. Li, et al., Impact of natural and social environmental factors on building energy

consumption: based on bibliometrics, J. Build. Eng. 37 (2021) 102136, https://doi.

org/10.1016/j.jobe.2020.102136.

[5] M. Hamed, S. Nada, Statistical analysis for economics of the energy development in

North Zone of Cairo, Int. J. Finance Econ. 5 (2019) 140–160.

[6] American Society of Heating, Refrigeration and Air-Conditioning Engineers,

ASHRAE handbook: Fundamentals, in: American Society of Heating, Refrigerating

and Air-Conditioning Engineers, ASHRAE, Atlanta, GA, USA, 2009.

[7] J. Runge, R. Zmeureanu, Forecasting energy use in buildings using artificial neural

networks: a review, Energies 12 (17) (2019) 3254, https://doi.org/10.3390/

en12173254.

Build. 40 (12) (2008) 2169–2176, https://doi.org/10.1016/j.enbuild.2008.06.013.

[8] Z. Dong, et al., Hourly energy consumption prediction of an office building based

on ensemble learning and energy consumption pattern classification, Energy Build.

241 (2021) 110929, https://doi.org/10.1016/j.enbuild.2021.110929.

[9] G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: a comparison of

regression analysis, decision tree and neural networks, Energy 32 (9) (2007)

1761–1768, https://doi.org/10.1016/j.energy.2006.11.010.

[10] Z. Yu, et al., A decision tree method for building energy demand modeling, Energy

Build. 42 (10) (2010) 1637–1646, https://doi.org/10.1016/j.enbuild.2010.04.006

[11] Hai-xiang Zhao, F. Magoul`es, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592, https://doi.org/10.1016/j.rser.2012.02.049.