J. Electrical Systems 20-1s (2024): 743-748

Kalpana D. Vidhate[1]*,

Dr. Pragya Nema[2]

Forecasting Energy Utilization in Residential Areas through Machine Learning Techniques

Journal of Electrical Systems

Abstract:
Analyzing and forecasting energy use in residential areas and buildings is critical for promoting sustainability and optimising energy usage. This paper describes a unique way for exact energy consumption analysis and forecasting that employs machine learning techniques such as linear regression and ridge regression. The proposed model intends to provide important insights and forecasts by combining historical energy data with relevant parameters such as building attributes, weather conditions, and tenant behaviour, improving energy usage in residential settings. The expected prototype is tested using a dataset accessible on Kaggle. The findings provided in the study show that our suggested model surpasses the state-of-the-art system, with accuracy rates of 66.56 percent for linear regression and 70.12 percent for ridge regression, respectively.

Keywords: *Regression, energy, Kaggle, machine learning, consumption, behavior, prototype, ridge, regression, forecasting, dataset, electricity, model construction.*

INTRODUCTION

Analyzing and predicting energy consumption are crucial elements in attaining energy efficiency and sustainability in residential areas and buildings. Precise forecasts of energy usage play a key role in optimizing management strategies, enhancing resource allocation, and fostering energy conservation practices. The incorporation of machine learning (ML) methods, such as linear regression (LR) and ridge regression (RR), has demonstrated promising outcomes, improving energy forecasting models in recent years. Numerous studies have delved into the realm of analyzing and predicting energy consumption in residential buildings through ML approaches.

Furthermore, the utilization of ridge regression has been investigated in the realm of energy consumption analysis. In a study by Researchers [2], ridge regression was utilized to scrutinize energy consumption patterns in residential areas, highlighting the importance of considering factors such as occupant behavior and energy management in predictive models. This paper introduces an innovative model for analyzing and predicting energy utilization in residential buildings, employing a combination of linear regression and ridge regression techniques to enhance accuracy. Building on this research, the proposed model extends beyond traditional linear regression by incorporating ridge regression, known for addressing multicollinearity.

It recognizes the influence of occupant behavior on energy patterns, providing a comprehensive framework. To refine predictions further, future research could explore integrating variables like building structure and environmental conditions, contributing to a more holistic understanding of energy usage patterns, and optimizing efficiency in residential buildings.

We aim to leverage historical energy data, analyze building attributes, consider weather conditions, and factor in occupant behavior to develop a thorough and accurate forecasting model. By incorporating a machine learning framework, our goal is to enhance the precision of energy consumption predictions, enabling informed decision-making in energy management.

The structure of the remaining sections in this article unfolds as follows. Section 2 provides an overview of pertinent work on the topic, delving into energy consumption analysis and forecasting utilizing machine learning. Section 3 meticulously delineates the proposed model, elucidating its key components. The experimental design and assessment findings are expounded upon in Section 4. Subsequently, Section 5 engages in a discussion of the outcomes, addressing limitations, and proposing potential directions for future research.

LITERATURE SURVEY:

During the evaluation phase, Williams and Gomez meticulously ensured the distinctiveness of training and testing sets,

[1]**Assistant Professor, Department of Electrical engineering, Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar.*
[2]*Professor, Department of Electrical engineering, Oriental University, Indore."*

maintaining exclusivity in both households and time. Their investigation yielded "residual mean squared errors (RMSE) for linear regression, regression trees, and MARS of 99.803 ± 3.057, 100.435 ± 3.441, and 94.286 ± 3.238 kBtu/day, respectively, for daily consumption projections. For monthly consumption estimates, the RMSE values were 25.743 ± 1.097, 18.277 ± 1.156, and 19.831 ± 1.187 kBtu/day for linear regression, regression trees, and MARS", respectively. This research convincingly showcases the effectiveness of machine learning methodologies, particularly regression trees, linear regression, and MARS, in accurately predicting energy consumption in residential areas. The study yields promising outcomes for understanding both daily and monthly consumption patterns.

The focus of Rodrigues et al.'s study is on estimating energy consumption in residential homes using an Artificial Neural Network (ANN) [5]. The collection, which includes hourly energy usage measurements from 93 residences in Lisbon, Portugal, spans six to eight months and has 93,744 entries. The researchers created two models, one for daily usage prediction and one for hourly usage prediction, using an ANN architecture with a single hidden layer of 20 neurons. The study's goal was to precisely anticipate household energy usage using machine learning techniques, producing trustworthy forecasts for both daily and hourly trends.

Williams and Gomez guaranteed the distinctness of training and testing sets during the assessment phase, ensuring exclusivity in both homes and time. In the evaluation of daily consumption estimates, the analysis produced residual mean squared errors (RMSE) of "99.803 ± 3.057, 100.435 ± 3.441, and 94.286 ± 3.238 kBtu/day for linear regression, regression trees, and MARS", respectively. For monthly consumption estimations, the corresponding RMSE values were "25.743 ± 1.097, 18.277 ± 1.156, and 19.831 ± 1.187 kBtu/day for linear regression, regression trees, and MARS". This investigation underscores the effectiveness of machine learning approaches, specifically regression trees, linear regression, and MARS, in accurately predicting energy consumption in residential areas. The outcomes offer promising insights into forecasting both daily and monthly consumption trends.

Shifting the focus to the research conducted by Rodrigues et al. [5], their study centers on predicting energy utilization in residential households using an Artificial Neural Network (ANN). Expanding on the significance of these findings, the successful demonstration of machine learning approaches, as evidenced in the studies, signifies a noteworthy advancement in energy consumption forecasting. The integration of regression trees, linear regression, and MARS proves to be instrumental in enhancing the accuracy of predictions for both short-term (daily) and long-term (monthly) energy usage trends.

The connection between these systems is facilitated through Keras, a TensorFlow API. The significance of smart city strategies is underscored in [7], emphasizing their role in mitigating CO_2 emissions and promoting energy conservation. The study advocates for decisive policies to enhance energy conservation efforts. In [8], a context-based methodology for estimating electricity use and output in buildings is proposed and tested. The exploration of various energy management systems (EMSs) is presented in [9], with a focus on addressing interactions among systems within a city's electric power infrastructure. The research presents a new idea: a distributed platform designed for evaluating collaborative EMS technology.

[10] developed a hybrid smart grid that uses solar (PV), hydro, and thermal energy sources to create power. The delivery system is optimised for real-time energy cost (ECRT), with delay concerns considered (FoL). The study's purpose is to provide resources for controlling power operations in smart grids more effectively. [11] examines the outcomes of a comprehensive framework for smart homes and smart cities, addressing technological and functional constraints.

RESEARCH METHODOLOGY

The research delves into the exploration of a prediction strategy employing supervised machine learning regression algorithms to estimate energy consumption in residential structures. Extracted from the Kaggle repository, the dataset encompasses information on residential structures across diverse cities. The prediction method will utilize linear regression and ridge regression machine learning techniques, implemented within a Jupyter notebook using the Python programming language [12].

Prior to model training and testing, the raw data undergoes meticulous analysis and pre-processing to address potential complications like missing data, ensuring the model's robustness during training. The research framework unfolds in four key sections: data analysis, where a comprehensive examination of the raw data transpires; model construction, where the linear regression and ridge regression models take shape based on insights gleaned from the data analysis; data visualizations, generating visual representations to enhance comprehension of patterns; and model evaluation, a critical step involving rigorous assessment of each model's performance to gauge the prediction strategy's effectiveness and accuracy. This systematic progression aims to establish a dependable framework for predicting energy consumption in residential structures, leveraging the capabilities of supervised machine learning regression algorithms.

DATA ANALYSIS:

Regression analysis serves as a powerful tool for gaining comprehensive insights that can be leveraged to further enhance goods and services. It enables a precise identification of the factors influencing a particular issue, distinguishing which elements are most significant, which ones can be disregarded, and understanding how these factors interact through regression analysis. The initial steps in this process involve identifying missing values, eliminating outliers, addressing skewness, and examining correlated variables. Subsequently, the model is constructed by applying both Linear and Ridge regression techniques.

Model Construction:

The proposed model incorporates both linear regression and ridge regression methodologies.

Linear Regression:

At the core of regression analysis lies linear regression, assuming a linear relationship between the dependent variable and predictor(s). The objective is to determine the best-fit line, illustrating the association between predictors and the dependent variable. Basic linear regression, a fundamental and widely used technique, predicts the outcome of a dependent variable using independent variables, establishing a linear relationship between them. Following linear regression, Ridge Regression is implemented.

Ridge Regression:

Tailored for the analysis of collinearity in multiple regression data, Ridge regression is a fundamental regularization technique, although its complexity deters some. While comprehending the underlying physics might seem challenging, grasping the general concept of multiple regression is crucial. The key distinction between regularization and regression lies in how model coefficients are calculated. A solid understanding of multicollinearity is essential for comprehending Ridge Regression.

L2 Regularization is performed using ridge regression. The coefficients in ridge regression are characterized as follows: for a predictor matrix $X \in R^\wedge(n\times p)$ and a response vector $y \in R^\wedge n$.

- The crucial parameter λ, determining the intensity of the penalty term, plays a significant role in this context.
- When $\lambda = 0$, the objective mirrors simple linear regression, resulting in identical coefficients as basic linear regression.
- For $\lambda = \infty$, the coefficients become 0 due to the infinite weight assigned to the square of the coefficients; any value below zero would render the objective infinite.
- In the range $0 < \lambda < \infty$, the magnitude of λ governs the emphasis placed on different components of the equation.
- In simpler terms, the minimized objective equals the Least Square (LS) Objective plus λ times the square root of the sum of coefficients.
- The LS Objective represents the linear regression objective without normalization.

Ridge regression in Rn introduces bias by decreasing coefficients in the direction of zero. However, it can significantly reduce variance, resulting in a lower mean-squared error. The ridge penalty is multiplied to control the amount of decrease. We can get different coefficient estimates for different values of since a higher value of causes more shrinkage.

Model Assessment:

Accuracy, Root Mean Squared Error (RMSE), Mean Squared Error (MSE), & Mean Absolute Error (MAE) were used to measure the model's estimating performance (MAE).

“Mean absolute error is a technique for determining the difference between the real or actual value and expected value (MAE). MAE scores around zero imply better performance, whereas values greater than zero suggest poor performance”.

The mean squared error is the assessment of a squared variance between both the estimated and actual values (MSE). The MSE formula is used to determine the quality of a prediction model. When the score is close to zero, MSE performance is good. The formula for Mean Absolute Error (MAE) is in eq. (1):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|y_j - \hat{y}_j\right| \tag{1}$$

Where:

- n is the number of observations.
- y_j represents the actual value for the j-th observation.
- $\hat{y}_j$ represents the predicted value for the j-th observation.

The Root Mean Squared Error is another way for evaluating the discrepancies between the estimated and actual values (RMSE). "It is calculated using the square root of the Mean Square Error (MSE)."

The formula for Root Mean Squared Error (RMSE) is in eq. (2):

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n} \left(y_j - \hat{y}_j\right)^2} \tag{2}$$

Where:

- n is the number of observations.
- y_j represents the actual value for the j-th observation.
- $\hat{y}_j$ represents the predicted value for the j-th observation.

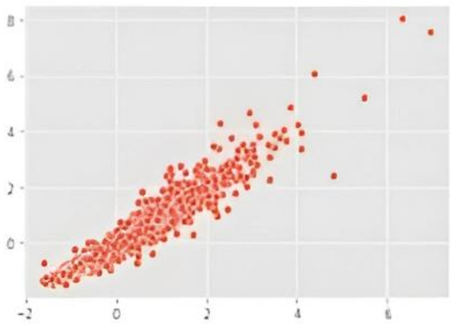
RESULT AND DISCUSSION:

The linear regression model was used to forecast energy use in residential areas and buildings. The model made use of a variety of input features, including building specifics (room size, year of construction, number of legrooms and headrooms, and so on) and meteorological data (humidity and temperature). The model's execution was assessed utilising the MAE, MSE, RMSE, and correctness as measurement metrics. Based on the experimental data, the linear regression model predicted energy consumption with reasonable accuracy.

Training Accuracy	63.51548
Testing Accuracy	66.56888
Mean Squared Error	0.6564
Mean Absolute Error	0.659
Root Mean Square Error	0.5602

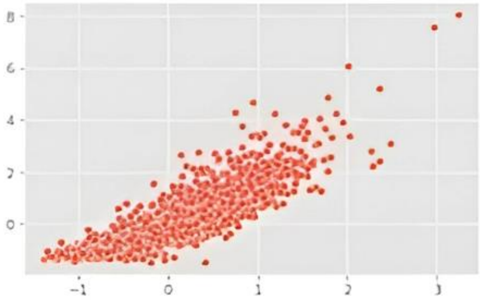
Following are the evaluation matrices for Linear Regression

Following is the scatter plot for linear regression.



Name	Score
|Training Accuracy	91.70448
Testing Accuracy	90.7518
Mean Squared Error	0.9711
Mean Absolute Error	0.998
Root Mean Square Error	0.9855

The scatter plot is seen in the image below.



Following are the evaluation matrices for Hybrid Regression

The study's results indicate that the linear regression model effectively captures the connections between input variables and energy consumption patterns in residential areas and structures.”Moreover, the investigation also employed the ridge regression model for predicting energy usage. Ridge regression, a regularization method, incorporates a penalty term to mitigate issues such as multicollinearity and overfitting. The ridge regression model utilized identical input attributes and assessment measures as the linear regression model.

Following are the evaluation matrices for Ridger Regressor-

Name	Score
Training Accuracy	68.08
Testing Accuracy	70.1244z
Mean Squared Brror	0.6821
Mean Absolute Error	0.7404
Root Mean Square Error	0.6259

The experimental findings showed that the ridge regression model predicted energy usage as well as the linear regression model. The Decision Stump serves as the default base estimator for the Gradient Boost method, maintaining a fixed structure. Analogous to AdaBoost, the estimator in the gradient boosting method can be customized. However, if the number of estimators (n_estimator) is not specified, the algorithm defaults to 100. This versatile method can operate as a regressor or classifier, accommodating predictions for both continuous and categorical target variables. When employed as a regressor, the cost function is Mean Square Error (MSE), while for classification tasks, it adopts Log Loss as the cost function. This adaptability makes the gradient boosting method a robust choice for a variety of predictive modeling scenarios.

CONCLUSION:

In summary, the research article centered on the analysis and forecasting of energy usage in residential areas and structures, employing linear regression and ridge regression models. The study successfully demonstrated the efficacy of these models in accurately predicting energy usage patterns. These findings have far-reaching implications for energy management, efficiency promotion, cost savings, and environmental stewardship. The study emphasises the importance of using machine learning algorithms in energy consumption analysis and forecasts, which provide vital insights for informed decision-making and the development of successful energy management methods.

REFERENCES:

1. Geoffrey K.F. Tso, Kelvin K.W. Yau,Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks,Energy, Volume 32, Issue 9, 2007, Pages 1761-1768,ISSN 0360-5442,
2. Tatiane C. Carneiro, Paulo A.C. Rocha, Paulo C.M. Carvalho, Luis M. Fernández- Ramírez,Ridge regression ensemble of machine learning models applied to solar and wind forecasting in Brazil and Spain, Applied Energy, Volume 314,2022,118936, ISSN 0306-2619.
3. H. xiang Zhao and F. Magoul`es, “A review on the prediction of buildingenergy consumption,” Renewable and Sustainable Energy Reviews,vol. 16, no. 6, pp. 3586 – 3592, 2012.
4. K. T. Williams and J. D. Gomez, “Predicting future monthly residentialenergy consumption using building characteristics and climate data: Astatistical learning approach,” Energy and Buildings, vol. 128, pp. 1 –11, 2016.

5. F. Rodrigues, C. Cardeira, and J. Calado, “The daily and hourly energyconsumption and load forecasting using artificial neural network method:A case study using a set of 93 households in portugal,” Energy Procedia,vol. 62, pp. 220 – 229, 2014.
6. Vázquez-Canteli, José R., et al. "Fusing TensorFlow with buildings energy simulation for intelligent energy management in smart cities." Sustainable cities and society 45 (2019): 243-257.
7. Wang, Stephen Jia, and Patrick Moriarty. "Energy savings from Smart Cities: A critical analysis." Energy Procedia 158 (2019): 3271-3276.
8. Gellert, Arpad, et al. "A study on forecasting electricity production and consumption in smart cities and factories." International Journal of Information Management 49 (2019): 546-556.
9. Fujimoto, Yu, Hideo Ishii, and Yasuhiro Hayashi. "Designing sustainable smart cities: cooperative energy management systems and applications." IEEJ Transactions on Electrical and Electronic Engineering 15.9 (2020): 1256-1270.
10. Khalil, Muhammad Ibrahim, et al. "Hybrid smart grid with sustainable energy efficient resources for smart cities." Sustainable Energy Technologies and Assessments 46 (2021): 101211.
11. Perera, H., & Costa, L. (2023, July 28). Personality Classification of Text Through Machine Learning and Deep Learning: A Review (2023). *International Journal for Research in Advanced Computer Science and Engineering*, *9*(4), 6–12. Https://Doi.Org/10.53555/Cse.V9i4.2266.
12. Kim, Hakpyeong, et al. "A systematic review of the smart energy conservation system: From smart homes to sustainable smart cities." Renewable and Sustainable Energy Reviews 140 (2021): 110755.