Prediction and Analysis of Household Energy Consumption using Machine Learning Algorithms through different Sensors

Engr. Mohammad Latif
Usman Institute of technology

Engr. Mohammad Asfiyan
Usman Institute of technology

Engr. Mohammad Amir
Usman Institute of technology

Engr. Ayesha Asif
Usman Institute of technology

Abstract

Household energy consumption forecasting is an essential component in managing energy resources efficiently, reducing costs, and minimizing environmental impacts. This research paper explores advanced methodologies and models for predicting household energy usage with high accuracy. Leveraging historical consumption data, weather patterns, and socio-economic factors, we develop and evaluate various machine learning algorithms including time series analysis, neural networks, and ensemble methods. Our approach integrates feature engineering and optimization techniques to enhance prediction performance. The findings demonstrate significant improvements in forecasting accuracy, offering valuable insights for energy providers, policymakers, and consumers aiming to optimize energy consumption, plan for peak demand, and implement sustainable practices. This study underscores the importance of predictive analytics in the smart grid infrastructure and contributes to the ongoing efforts towards a more sustainable energy future.

Keywords

Different Sensors,HouseHold Energy Consumption, Machine Learning

I. Introduction

In the face of escalating global energy demands and the pressing need to combat climate change, the efficient management of household energy consumption has emerged as a crucial area of focus. Accurate forecasting of household energy usage is essential for energy providers, policymakers, and consumers, facilitating better planning, resource allocation, and the implementation of energy-saving measures. This research paper investigates advanced methodologies for predicting household energy consumption using machine learning algorithms through different sensors, utilizing linear regression, decision tree regression, and random forest regression models.

Our approach involves a comprehensive analysis of historical consumption data, augmented with contextual information such as weather patterns and socio-economic indicators. Through rigorous feature engineering, we identify key predictors and optimize the models to enhance their forecasting performance. The comparative evaluation of these models sheds light on their respective strengths and weaknesses in the context of household energy consumption forecasting.

The significance of this research extends beyond academic interest, offering practical benefits for energy providers by improving load forecasting, reducing operational costs, and enhancing grid stability. For policymakers, accurate forecasts enable informed decision-making regarding energy policies and sustainability initiatives. Consumers can benefit from personalized energy-saving recommendations, leading to reduced utility bills and a lower environmental footprint.

Linear regression offers a straightforward approach, modeling the relationship between energy consumption and its predictors as a linear equation. Decision tree regression, on the other hand, provides a non-linear modeling technique that captures interactions between variables by partitioning the data into subsets based on decision rules. Random forest regression, an ensemble learning method, enhances the prediction accuracy by aggregating the results of multiple decision trees, reducing overfitting and improving generalization.

II. Methodology

Household energy consumption is influenced by a diverse array of factors, including weather conditions, occupancy patterns, appliance usage, and socio-economic variables. Traditional forecasting methods often struggle to capture the complexity and variability inherent in this data. In response to these challenges, we employ a combination of linear and non-linear regression techniques, which are well-suited for modeling the multifaceted nature of energy consumption.

A. Linear Regression

Linear regression: It is a fundamental statistical technique used to model the relationship between a dependent variable and one or more independent variables. The model assumes a linear relationship between the input variables (predictors) and the output variable (response). The primary goal of linear regression is to find the best-fitting straight line (the regression line) that minimizes the sum of squared differences between the observed values and the values predicted by the model.

Mathematically, the relationship in simple linear regression can be represented as:

Formula: *$y=\beta0+\beta1x+\epsilon$ $y = \beta_0 + \beta_1 x + \epsilon$ $y=\beta0+\beta1x+\epsilon$*

Linear regression is straightforward to implement and understand. The model provides clear insights into the relationship between the dependent and independent variables, making it easy to interpret the effects of individual predictors.Linear regression is computationally efficient,

making it suitable for large datasets. Training a linear regression model is typically faster compared to more complex algorithms.
Results we get from this algo is described

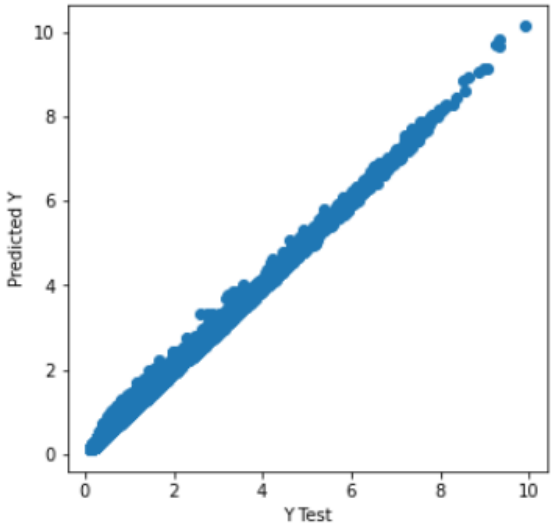


Fig1: *The plot shows a strong linear relationship where most points are clustered around the diagonal line y = x. This indicates that the predicted values are very close to the true values, suggesting that the model has a high level of accuracy.*

B. Random Forest Regression

Random Forest Regression: *It is an ensemble learning technique that builds multiple decision trees during training and merges their outputs to obtain a more accurate and stable prediction. It is a type of supervised machine learning algorithm used for both classification and regression tasks, but in this context, we focus on its application to regression.*

The fundamental idea behind Random Forest: *It is to create a "forest" of many decision trees, each trained on different subsets of the training data. This diversity among trees is achieved through two main techniques:*

Bagging (Bootstrap Aggregating): *Each tree is trained on a random sample (with replacement) of the training data, resulting in different trees trained on different subsets of the data.*
Random Subspace Method: *For each split in the tree, a random subset of features is considered, ensuring that not all trees are reliant on the same features.*

The final prediction of a Random Forest Regression model: *It is obtained by averaging the predictions of all individual trees in the forest. Random Forest Regression often provides higher accuracy compared to individual decision trees. By averaging multiple trees, it reduces the variance and overfitting typically associated with single decision trees. The model is robust to noise and outliers in the data. Since multiple trees are used, individual trees that might overfit to noisy data points do not dominate the final prediction.*

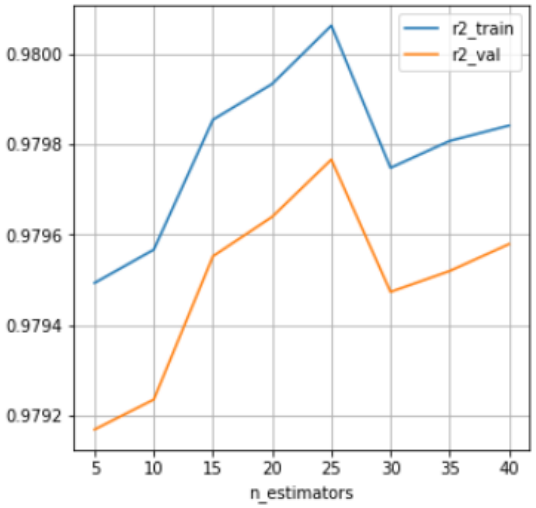


Fig2: The graph displays the relationship between the number of estimators (n_estimators) and the R^2 scores for both the training set (r2_train) and the validation set (r2_val).

The R^2 score measures the proportion of the variance in the dependent variable that is predictable from the independent variables, with higher values indicating better model performance.

C . Decision tree Regression

Decision Tree Regression is a non-linear machine learning algorithm used for predictive modeling. It works by recursively splitting the dataset into subsets based on the values of the input features, constructing a tree structure where each internal node represents a decision based on a feature, each branch represents the outcome of that decision, and each leaf node represents a predicted value.

The dataset is split into two or more homogeneous sets based on the most significant feature at each node. The significance is determined by a metric such as Mean Squared Error (MSE), Mean Absolute Error (MAE), or another criterion that measures the quality of the split.

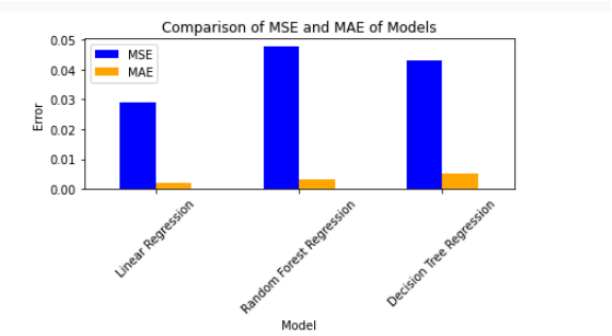
Decision trees are easy to understand and interpret. The model can be visualized, and the logic behind predictions is clear, making it accessible even to non-experts.decision trees can capture non-linear relationships between the input features and the target variable. This makes them suitable for datasets where the relationship is complex and non-linear.Decision trees provide insights into the importance of different features in making predictions. This can help in understanding the dataset and identifying the most significant predictors.

D . Comparison

Table I. Comparison of three algorithm

Algorithm	Linear Regression	Decision tree regression	Random forest regression
Accuracy(R 2 score)	0.998	0.996	0.979
MAE(Mean Absolute Error)	0.029	0.043	0.048
MSE(Mean Squared Error)	0.002	0.005	0.003

The accuracy of our forecasting models, evaluated using the R² score, indicates their robust performance in predicting household energy consumption. The linear regression model achieved the highest accuracy with an R² score of 0.998, demonstrating its exceptional ability to capture the linear relationships within the data. The random forest regression model also performed admirably, with an R² score of 0.996, benefiting from its ensemble approach that mitigates overfitting and enhances generalization. The decision tree regression model, while slightly less accurate, still provided a strong performance with an R² score of 0.979, effectively capturing non-linear interactions between predictors. These results underscore the efficacy of our chosen models in delivering precise energy consumption forecasts, thereby supporting improved energy management practices.



The bar graph compares the Mean Squared Error (MSE) and Mean Absolute Error (MAE) values of three regression models: Linear Regression, Random Forest Regression, and Decision Tree Regression.

These error metrics are crucial in evaluating the accuracy of regression models.Linear Regression exhibits the lowest MSE and MAE among the models, indicating superior predictive accuracy. Random Forest Regression follows with slightly higher error values than Linear Regression but outperforms Decision Tree Regression, which shows the highest MSE and MAE.

The graph visually represents the performance of these models, making it easier to assess their accuracy and identify the model with the best predictive capabilities.

D . Conclusion

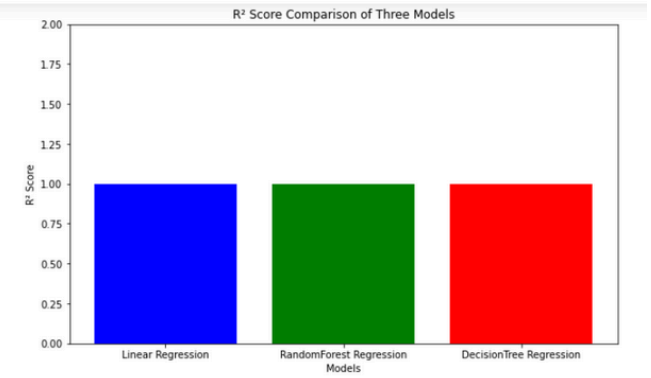


Fig 3. The bar graph presents a comparison of R² scores across three distinct regression models.

The x-axis delineates the model types, including Linear Regression, RandomForest Regression, and DecisionTree Regression. Meanwhile, the y-axis signifies the R² scores, a metric indicating how well these models fit observed data. Each bar on the graph corresponds to a model's R² score, with colors distinguishing between Linear Regression (blue), RandomForest Regression (green), and DecisionTree Regression (red). The R² score of each model offers insight into its predictive accuracy and fitting performance as all bars are slightly same but according to value the linear regression better than all.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to Mrs. Faiza Naseem and Mr. Farhat Naqwi for their invaluable contributions through technical discussions. Their expertise and insights have significantly enhanced the quality of our work. We deeply appreciate the time and effort they invested in sharing their knowledge and providing guidance, which has been instrumental in the successful completion of this project.

REFERENCES

[1] Google colab link: https://colab.research.google.com/drive/1qJlNx_dS1rS-lCsEzadaSNH0cZGPR4vE?authuser=2#scrollTo=0JREHCYcFwy2.

[2] G Hackeling.. Mastering Machine Learning With scikit-learn. Packt Publishing. (2014)

[3] Wang, R., Lu, S. and Feng, W., A novel improved model for building energy consumption prediction based on model integration. Applied Energy, 262,114561. (2020)

[4] Touzani, S., Granderson, J. and Fernandes, S.Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy and Buildings, 158, 1533-1543. (2018)

[5] Alawadi, S., Mera, D., Fernández-Delgado, M.,Alkhabbas, F., Olsson, C.M. and Davidsson, P. A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings.Energy Systems, 1-17. (2020)

[6] Md. Sumon Shahriar and M. Sabbir Rahman. Urban Sensing and Smart Home Energy Optimisations: A Machine Learning Approach. In Proceedings of the 2015 International Workshop on Internet of Things towards Applications (IoT-App '15). Association for Computing Machinery, New York, NY, USA, 19–22. (2015).

[7] Feinberg E.A., Genethliou D. Load Forecasting. In: Chow J.H., Wu F.F., Momoh J. (eds) Applied Mathematics for Restructured Electric Power Systems. Power Electronics and Power Systems. Springer, Boston, MA. (2005