JOURNAL OF
EMERGING INVESTIGATORS

Article

The influence of economic factors on United States household energy consumption in 2020

Aditya Ramanathan[1], Tristan Ballard[2]
[1] Harker School, San Jose, California
[2] Sust Global, San Francisco, California

SUMMARY

Household energy consumption is a major contributor to both greenhouse gas emissions and overall energy demand. Understanding what factors most influence household energy use is crucial to help policymakers develop more effective energy conservation and pricing strategies. We hypothesized that in 2020, economic factors, specifically the cost of natural gas and electricity, were stronger predictors of U.S. household energy consumption than traditional physical and environmental factors like residence size and climate. To test this hypothesis, we analyzed data from 18,496 households in the 2020 Residential Energy Consumption Survey. We applied several machine learning models including random forest, gradient boosting, and polynomial regression to identify the most significant factors contributing to residential energy use. The gradient boosting model performed the best, explaining 91.1% of the variation in household energy consumption. A detailed analysis of the model's feature importance supported our hypothesis, showing that the cost of natural gas and electricity were the dominant factors influencing energy consumption. These results suggest that energy costs were strongly associated with household energy consumption in 2020. This pattern may reflect increased sensitivity to energy prices during a year marked by economic uncertainty related to the COVID-19 pandemic.

INTRODUCTION

Understanding household energy consumption is critical for developing effective energy policies and reducing the environmental impact of energy generation, a major contributor to greenhouse gas emissions. As research continues to identify new sustainable energy practices, evaluating shifts in energy trends is essential to focus sustainability efforts on areas of higher energy consumption. In the United States (U.S.), household energy consumption represents 20% of the national carbon footprint (1). On average, 17,302 pounds of carbon dioxide were released by each household in 2019 (2). Although the U.S. doubled its renewable energy deployment in 2020 compared to 2019, the nation needs to triple its sustainable energy deployment to meet the Biden administration's clean energy targets (3).

Multiple determinants contribute to household energy consumption, including size of the house, region, climate, and appliance usage. The recent datasets from energy surveys provide opportunities to study these determinants and how they impact energy use (4, 5). Machine learning offers a powerful method for analyzing complex and extensive datasets from energy surveys, offering a more efficient alternative to cumbersome manual analysis. Several studies have used machine learning and deep learning using neural networks to study energy consumption in various contexts (6-8). Predictive machine learning models have been developed to forecast household energy consumption (9, 10). However, these studies are often limited by datasets that are not generalizable due to their narrow geographic focus or smaller sample sizes (10).

Several datasets track energy consumption, including those from the National Renewable Energy Laboratory (NREL), the American Council for an Energy-Efficient Economy (ACEEE), and the U.S. Energy Information Administration (EIA). Due to its comprehensive nature, the Residential Energy Consumption Survey (RECS) dataset is used in many studies. Previous research using RECS data has primarily focused on identifying key factors driving energy consumption and developing predictive models for specific time periods or geographic regions (11-14). However, these studies have not explored how recent changes in household characteristics and energy technologies affect consumption patterns in the 2020 dataset.

This study addresses limitations in previous research related to outdated and non-generalizable data by using the most recent Residential Energy Consumption Survey (RECS) dataset, released in March 2024, which captures household energy data from 2020. This dataset is notable because 2020 was marked by the COVID-19 pandemic, which forced widespread shelter-in-place measures and economic uncertainty. While many factors can influence household energy use, we hypothesized that in 2020, economic factors, such as the cost of natural gas and electricity, had a greater influence on U.S. household energy consumption than traditional physical and environmental factors, including house size, building age, and climate. To test our hypothesis, we applied a suite of machine learning algorithms to the RECS dataset to identify the model that predicts energy consumption most accurately. Building on prior work using elastic net regression and gradient boosting models, this study applies similar machine learning methods to the most recent RECS data to provide updated insights into residential energy consumption (14-16). Then, we extracted the key factors from the model and verified our hypothesis. This understanding of how household energy consumption changes over time can reveal changes in consumption patterns at local and national levels and therefore inform effective policies and future planning for sustainable energy use.

JOURNAL OF
EMERGING INVESTIGATORS
https://doi.org/10.59720/25-239

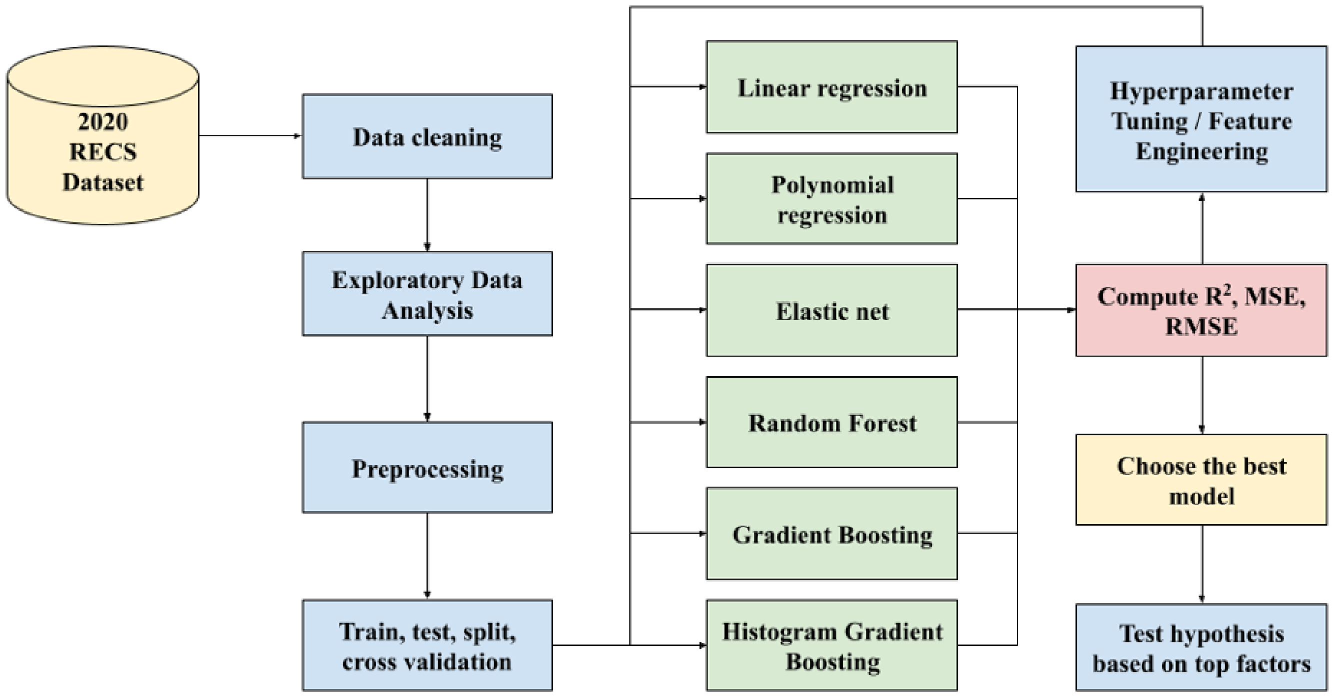

Figure 1: Machine learning methodology pipeline. This flowchart outlines the methodology used to analyze the 2020 RECS dataset and develop a predictive model for household energy consumption. The process began with data preparation, preprocessing, model exploration, feature engineering, and proceeded to the evaluation of six different machine learning models. The best-performing model was selected based on key performance metrics like R-squared (R^2), mean squared error (MSE), and root mean squared error (RMSE) and used to test the hypothesis. Made with Google Drawings.

RESULTS

The 2020 RECS dataset captures a broad range of household characteristics, including occupant count, square footage, year of construction, appliance usage, energy spending by source, and geographic information such as region and climate type, providing a comprehensive foundation for modeling residential energy consumption (**Table A1**).

To identify the key determinants of residential energy consumption and test our hypothesis that economic factors were the dominant drivers in 2020, we trained and evaluated multiple machine learning models on the RECS dataset (**Figure 1**). We compared several algorithms to determine which model best captured the complex relationships within household energy data while also providing reliable feature importance rankings. Before evaluating model performance, we examined the distribution of energy sources among U.S. households in the dataset. Electricity and natural gas were the primary energy sources for most households, confirming that these fuels represent the largest contributors of residential energy use in 2020 (**Figure 2**).

We evaluated the models using the coefficient of determination, mean squared error, and root mean squared error. The main metric used was the coefficient of determination (R^2) which measures how closely the model fits the data. We also computed the root mean squared error, the average magnitude of the error between predicted values and actual observed values in a model (RMSE), and the mean squared error (MSE), the square of the RMSE (**Table 1**). Prediction errors increased as the total energy consumption increased. To reduce the influence of extreme values, 6 (or 0.03% of total) households with total energy consumption greater than 500 million British Thermal Units (BTUs) were excluded from the analysis.

Among all tested models, the gradient boosting model was the best performer, achieving an R^2 value of 0.911 (**Table 1**). This indicates that the model explained 91.1% of the variability in household energy consumption in the test dataset. The RMSE was 14,918, meaning the model's predictions were typically off by roughly 3% of the observed energy range of 1,182 to 506,714 (x 1,000 BTUs), suggesting reasonable overall accuracy (**Table 1**). Predicted values generally aligned with observed values (**Figure 3**). Therefore, to determine which factors most influenced household energy consumption, we used feature importance analysis from our gradient boosting model.

The most dominant factors influencing energy consumption in 2020 were the cost of natural gas and the cost of electricity, with feature importance magnitudes of 60.1% and 13.4% respectively (**Figure 4**). This was followed by heating degree days and house size with feature importance magnitudes of 5.9% and 5.1% respectively (**Figure 4**). Heating and cooling degree days estimate the energy needed for temperature regulation, reflecting the impact of climate conditions on energy consumption. Regional variation was also observed; while region was not among the top predictors overall, New England showed the highest feature importance (1.7%) among all regions (**Figure 4**).

DISCUSSION

This study investigated how household characteristics, energy costs, and climate factors predict residential energy

https://doi.org/10.59720/25-239

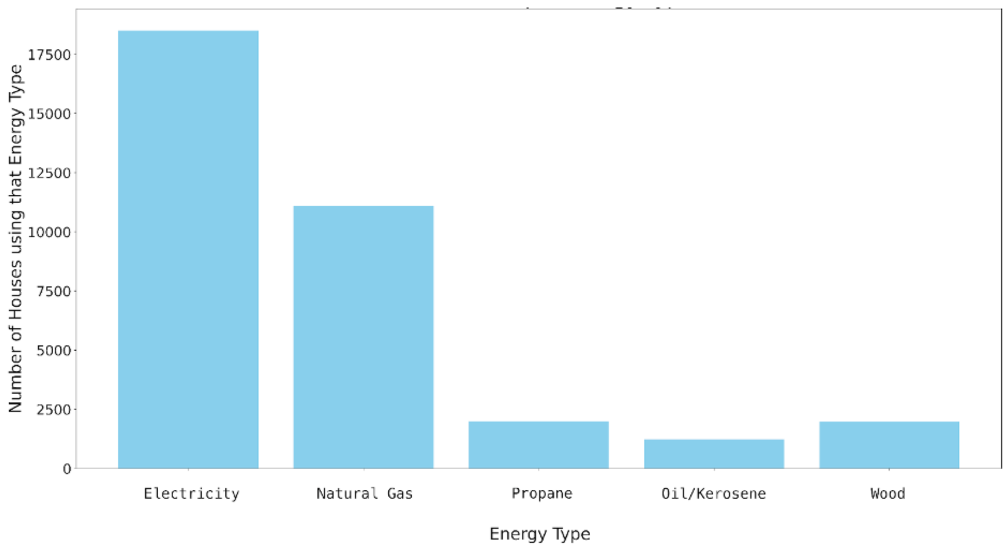

Figure 2: House count per energy source. This bar chart shows the number of houses versus the energy source used in the 2020 RECS dataset.

consumption in the U.S. using machine learning models applied to the 2020 RECS dataset. We hypothesized that energy cost variables would emerge as strong predictors. Among the models tested, gradient boosting performed best with the cost of natural gas, cost of electricity, and heating degree days emerging as the most important predictors. These findings suggest that price sensitivity and climate-related factors were central drivers of residential energy consumption, with meaningful implications for national energy policy.

By modeling and analyzing these results, the machine learning models effectively captured household energy consumption patterns in the U.S. The best-performing model, gradient boosting, explained 91.1% of the variability in a held-out test set. This strong performance suggests that the model effectively captured complex relationships within the RECS dataset.

Gradient boosting's strong performance is well suited to the RECS dataset, which has high dimensionality, feature complexity, feature interdependencies, non-linear relationships, and potentially noisy data. Examining the results across multiple tree-based models including random forest, gradient boosting, and histogram gradient boosting for the 2020 RECS dataset reveals differing feature rankings across the models, with the cost of natural gas, the cost of electricity, and the heating degree days identified as the most important by the top-performing model, gradient boosting. The tree-based models demonstrated the strongest performance, and they consistently showed that the cost of natural gas, the cost of electricity, and the heating degree days are the most important factors. This consistency strengthens our confidence that these three factors are the most reliable way to predict residential energy consumption. These results support our hypothesis: the cost of natural gas and electricity were indeed among the strongest predictors of household energy consumption. The prominence of heating degree days suggests that climate-related factors played a significant role, reflecting the evolving influence of climate on residential energy use.

This study found the cost of natural gas, the cost of electricity, and the heating degree days to be more influential predictors, in contrast to prior research by Burnett and Kiesling, which identified square footage, gas prices, and the number of bedrooms as primary predictors (11). This study also found that both gradient boosting and histogram gradient boosting outperformed random forest, in contrast to the prior work where random forest was the top performer (11). These differences may reflect both the expanded scale of the 2020 RECS dataset compared to earlier studies, as well as advances in ensemble modeling techniques that have become more accessible in modern software.

The prominence of energy prices as predictors suggests that households demonstrated increased price sensitivity in 2020, with most people sheltering in place due to the pandemic. The uncertainty around employment caused by the pandemic-related economic shutdown may have resulted in households being careful with their expenses (17). Although this study cannot establish causation, the results indicate an association between energy prices and consumption patterns. Heating and cooling degree days also emerged

Model	R^2	MSE	RMSE
Linear regression	0.828	431,921,053	20,782
Quadratic regression	0.894	264,553,219	16,265
Elastic net	0.828	431,994,709	20,784
Random forest	0.889	277,661,288	16,663
Gradient boosting	0.911	222,556,768	14,918
Histogram gradient boosting	0.905	237,485,087	15,410

Table 1: Model Performance Metrics. This table presents the performance metrics for the six machine learning models used to predict energy consumption. The models were evaluated using R-squared (R^2), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

https://doi.org/10.59720/25-239

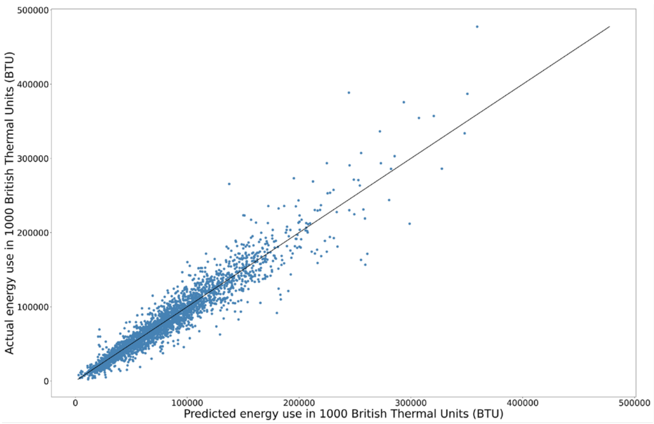

Figure 3: Predicted vs. actual energy use for gradient boosting model. This scatter plot compares the predicted energy use values from the gradient boosting model against the actual energy use values from the held-out dataset.

as significant predictors, highlighting the continued role of climate in shaping residential energy demand.

Interestingly, natural gas prices were highly predictive even for households that do not consume natural gas directly. Natural gas prices are strongly correlated with electricity prices, as natural gas fuels approximately 40% of U.S. electricity generation (19). Thus, natural gas prices capture overall energy market conditions that affect all households, whether they use natural gas directly or consume it indirectly through electricity.

Electricity remains the dominant household energy source, and a substantial portion of U.S. electricity generation relies on fossil fuels (19). Burning these fuels releases both particulate matter and greenhouse gases into the atmosphere. Additionally, 18.9% of electricity comes from nuclear fission, which generates radioactive waste requiring safe disposal (19). Policies that improve energy efficiency can help drastically reduce greenhouse gas emissions by fostering reductions in household energy consumption.

Several limitations of this study warrant consideration. First, while the 2020 RECS dataset is the most comprehensive survey of U.S. household energy use available, it only represents about 0.015% of all U.S. households (20). Second, the decision to exclude households using propane or wood may reduce generalizability to rural communities that may rely on these fuels. Finally, 2020 was an atypical year due to the COVID-19 pandemic, which could have substantially altered household behavior and energy use patterns. Future datasets from subsequent RECS surveys will be valuable for determining whether the patterns identified here persist over time. Future research could explore deep learning models to improve the accuracy of predicting household energy consumption. Climate-specific or state-level models may better capture localized energy patterns. Fumo and Biswas proposed that future research should consider building models tailored to individual dwellings (12). A personalized energy dashboard built on these methods could help homeowners track consumption and improve efficiency at scale. Future RECS datasets could be used to evaluate whether policy changes or market shifts alter the relative importance of economic and environmental factors over time.

Overall, these results suggest that economic factors, particularly energy costs, were the dominant predictors of U.S. residential energy consumption in 2020, outweighing traditional variables like house size and climate. These findings underscore the importance of energy pricing policy as a lever for influencing household consumption.

MATERIALS AND METHODS

Data source

The data used in this study were obtained from the 2020 Residential Energy Consumption Survey (RECS). RECS is the official survey collected by the EIA that collects detailed information about energy use in homes (20). The survey is collected every four to five years, and its purpose is to gather data on household energy consumption patterns, characteristics, and efficiency. The RECS dataset has high dimensionality and includes each household's characteristics, such as the number of occupants, square footage, the year of construction, and the number of rooms, as well as the types and amounts of energy consumed, including electricity, natural gas, propane, oil, and wood. The dataset also contains the various types of appliances used by the household, such as heating and cooling systems, lighting, and electronics. The survey also contains how much households are spending on energy, broken down by energy source and other relevant factors. Finally, it covers geographic information such as states, regions, weather, and climate types.

The 2020 RECS dataset (released in March 2024) includes information from 18,496 U.S. households. This dataset contains various features, socioeconomic factors, and climate zones, and contains the five types of energy consumption noted above. Energy consumption in the 2020 RECS microdata is reported in thousand British Thermal Units (1,000 BTUs), broken down by fuel type: electricity, natural gas, propane, oil/kerosene, and wood.

Data feature selection

The 2020 RECS dataset contains many factors for several households around the nation. However, some houses do not have certain features, such as swimming pools or electric cars. All partially populated columns meeting such criteria were removed to improve model generalizability. Other data simplification changes were made as well. Columns such as the number of cell phones, laptops, tablets, and printers were condensed into one column containing the number of electronic devices. The data on whether the tenant rents a house, with or without paying rent, was grouped under a single column indicating whether the house was rented or not. Regarding energy consumption, this dataset contains the amounts (in 1,000 BTU units) of electricity, natural gas, propane, oil/kerosene, and wood consumption. Most households do not consume propane, oil/kerosene, or wood. Fewer than 2,500 out of 18,496 houses use each of these energy sources **(Figure 2)**. Although this may reduce representation of rural households using alternative fuels, electricity and natural gas together constitute the primary energy sources for the vast majority of U.S. homes, ensuring the findings remain generalizable to mainstream residential energy consumption patterns. Based on this, the households that use these energy sources were removed, a total of 4,206 households, to avoid biasing the next steps in modeling and data analysis. The study conducted by Burnett and Kiesling

https://doi.org/10.59720/25-239

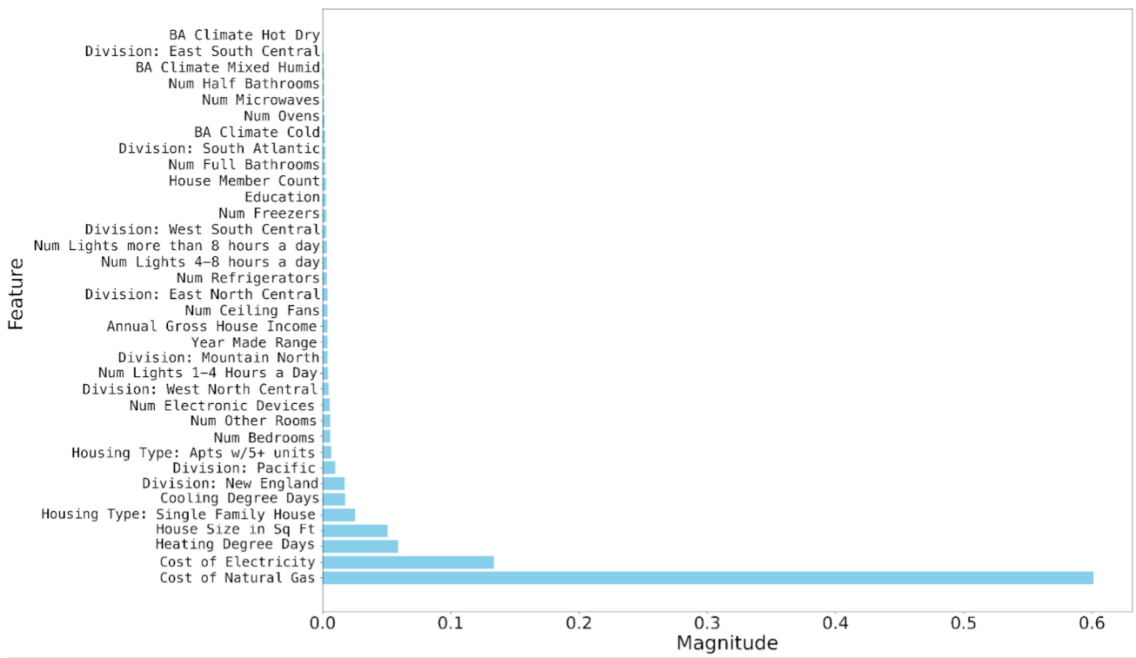

Figure 4: Gradient boosting feature contribution. This bar chart illustrates the relative contribution of different features to the gradient boosting model's predictions. The magnitude on the x-axis represents the importance of each feature.

used the total energy consumption in 1,000 BTU units provided in the RECS data from the years past (11). Building on the success of this earlier study, a total energy column was created by summing the total electricity consumption and total natural gas consumption, the only remaining energy sources after the previous step.

Summary statistics

Summary statistics for the key household features in the 2020 RECS dataset, including energy costs, house size, occupancy, and appliance counts, are presented **(Table A1)**. The minimum cost of electricity being negative is a valid data point as it represents households generating more electricity through solar panels than they consume. Heating and cooling degree days are measures that indicate the demand for heating or cooling. Thus, the colder the day, the higher the value of heating degree days. The summary statistics associated with individual features vary significantly relative to their central tendency. This represents the heterogeneity of the households that contributed data to the RECS dataset.

Data preprocessing

The initial categorical and ordinal data had been encoded using strictly increasing numbers. Since numerical ordering causes a natural hierarchy in values that do not have a hierarchy, such as categorical non-ordinal data, the data encoding needed to be reverted (23). The raw categorical data were used as labels for the charts for better readability. One-hot encoding was used to re-encode the categorical data, as this encoding method can handle categorical variables without imposing a natural ordering. In addition, the model can more easily handle new input values and is more efficient in terms of memory and computation (23).

Six households with a total energy consumption greater than 500,000,000 BTUs were removed as outliers. These changes reduce data leakage and improve feature quality for machine learning, benefiting both model performance and speed of computation. Lastly, machine learning includes training on a portion of the dataset and testing the model using the remainder of the dataset. A common ratio is to split the data into 80% training data and 20% testing data (24). This was the approach taken in this study.

Machine learning models

To ensure the best possible results were achieved, various machine learning models that worked for similar research were incorporated into this study: linear regression, polynomial regression, elastic net regression, random forest, and gradient boosting.

The scikit-learn Python library was used to program and train the machine learning models. A linear regression model was trained as a baseline to compare against the results from machine learning models. Basic quadratic regression, cubic regression, elastic net regression, random forest, gradient boosting, and histogram gradient boosting without any hyperparameter tuning were then run to see how they behaved relative to the linear regression model. The random forest model, elastic net model, and gradient boosting models performed marginally better with default parameters. The performance of these models improved significantly by running comprehensive hyperparameter tuning.

For hyperparameter tuning, the scikit-learn Grid Search function was employed to automate various permutations of parameters to determine which yielded the best results on the training data. For all the trials with Grid Search on the various machine learning methods, 10-fold cross-validation was used. This ensured that the model was not overfitting to the training data and would generalize well on the testing data for high accuracy. In addition, as described earlier, the removal of correlated columns prevents data leakage. The model performance was evaluated primarily using the R^2 metric, which provides a measure of how well the model explains the variance in the target variable. This metric was chosen because of its effectiveness in comparing model performance. The mean squared error (MSE) and root mean squared error (RMSE) were computed as well.

Feature engineering

While training the models and obtaining results, the following features were removed from the model. In some cases, these features were not only unlikely to have contributed to energy consumption, but also likely to have introduced noise, affecting the effectiveness of the model. The following features were removed: householder race, employment status, rent or own, Hispanic/Latino, number of weekdays spent at home, housing type: mobile home, housing type: single-family attached, housing type: apartment in building with 2-4 units, division: mountain south, climate: hot-humid, climate: marine, climate: mixed-dry, climate: very cold. Columns whose coefficients were insignificant in determining household energy, such as these, were removed. Once the columns were removed, the models were recomputed and error metrics were verified. This was done iteratively while verifying that the error metrics improved or remained constant.

Received: August 17, 2025
Accepted: February 4, 2026
Published: June 8, 2026

REFERENCES

1. Goldstein, Benjamin, et al. “The carbon footprint of household energy use in the United States.” *Proceedings of the National Academy of Sciences*, 117(32), 19122-19130, 20 Jul. 2020. doi.org/10.1073/pnas.1922205117
2. “Energy data facts. (n.d.). Residential Program Guide.” rpsc.energy.gov/energy-data-facts. Accessed: April 2024
3. McLaughlin, Katrina, and Lori Bird. “The US set a record for renewables in 2020, but more is needed”, *World Resources Institute*. www.wri.org/insights/renewable-energy-2020-record-us. Accessed: Jul 2024
4. Ushakova, Anastasia, and Slava Jankin Mikhaylov. “Big data to the rescue? Challenges in analyzing granular household electricity consumption in the United Kingdom.” *Energy Research & Social Science*, 64, 101428, Jun. 2020. doi.org/10.1016/j.erss.2020.101428
5. Zhou, Kaile, and Shanlin Yang. “Understanding household energy consumption behavior: The contribution of energy big data analytics,” *Renewable and Sustainable Energy Reviews*, 56, 810-819, Apr. 2016. doi.org/10.1016/j.rser.2015.12.001
6. Reddy, Sujan, et al. “Stacking deep learning and machine learning models for short-term energy consumption forecasting.” *Advanced Engineering Informatics*, 52, 101542, Apr. 2022. doi.org/10.1016/j.aei.2022.101542
7. Culaba, Alvin B., et al. “Machine learning-based energy consumption clustering and forecasting for mixed-use buildings.” *International Journal of Energy Research*, 44(12), 9659-9673, 18 May 2020. doi.org/10.1002/er.5523
8. Torabi, Mehrnoosh, et al. “A hybrid clustering and classification technique for forecasting short-term energy consumption.” *Environmental Progress & Sustainable Energy*, 38(1), 66-76, 27 Jun. 2018. doi.org/10.1002/ep.12934
9. Chou, Jui-Sheng., and Duc-Son Tran. “Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders.” *Energy*, 165, 15 Dec. 2018, pp. 709-726. doi.org/10.1016/j.energy.2018.09.144
10. Krishnan, Lokesh, et al. “Prediction of electrical power consumption in the household: Fresh evidence from machine learning approach.” *Energy Efficiency*, 16(7), 22 Sep 2023. doi.org/10.1007/s12053-023-10155-z
11. Burnett, J. Wesley, and L. Lynne Kiesling. “How do machines predict energy use? Comparing machine learning approaches for modeling household energy demand in the United States.” *Energy Research & Social Science*, 91, 102715, Sep. 2022. doi.org/10.1016/j.erss.2022.102715
12. Fumo, Nelson, and M. A. Rafe Biswas. “Regression analysis for prediction of residential energy consumption.” *Renewable and Sustainable Energy Reviews*, 47, 332-343, Jul. 2015. doi.org/10.1016/j.rser.2015.03.035
13. Tso, Geoffrey, and Jinging Guan. “A multilevel regression approach to understand effects of environment indicators and household features on residential energy consumption.” *Energy*, 66, 722-731. 01 Mar. 2014. doi.org/10.1016/j.energy.2014.01.056
14. Zhang, Wenwen, et al. “Estimating residential energy consumption in metropolitan areas: A microsimulation approach.” *Energy*, 155, 162-173, 15 Jul. 2018. doi.org/10.1016/j.energy.2018.04.161
15. Cui, Xue, et al. “Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: Toward energy-efficient buildings.” *Energy and Buildings*, 309, 113997, 15 Apr. 2024. doi.org/10.1016/j.enbuild.2024.113997
16. Friedman, Jerome H. “Greedy function approximation: A gradient boosting machine.” *The Annals of Statistics*, 29(5), Oct. 2001. doi.org/10.1214/aos/1013203451
17. “Unemployment rises in 2020, as the country battles the COVID-19 pandemic: monthly labor review.” *U.S. Bureau of Labor Statistics*. www.bls.gov/opub/mlr/2021/article/unemployment-rises-in-2020-as-the-country-battles-the-covid-19-pandemic.htm. Accessed: Aug. 2024.
18. Das, Debadyuti, et al. “Impact of COVID-19 on changing consumer behavior: lessons from an emerging economy.” *International journal of consumer studies*, 46(3), 692-715, 04 Feb. 2022. doi.org/10.1111/ijcs.12786
19. “Frequently Asked Questions (FAQs) (n.d.).” *U.S. Energy Information Administration (EIA)*. www.eia.gov/tools/faqs/faq.php?id=427&t=3. Accessed: Apr. 2024.
20. “Residential energy consumption survey. (n.d.).” *U.S. Bureau of Labor Statistics*. www.bls.gov/cex/cecomparison/recs_profile.htm. Accessed: Jul. 2024.
21. “Residential Energy Consumption Survey (RECS).” *U.S. Energy Information Administration (EIA)*. www.eia.gov/consumption/residential/data/2020/index.php?view=microdata. Accessed: Apr. 2024.
22. Kim, YouJeong, et al. “Comparison of linear and nonlinear statistical models for analyzing determinants of residential energy consumption.” *Energy and Buildings*, 223, 110226, 15 Sep 2020. doi.org/10.1016/j.enbuild.2020.110226

https://doi.org/10.59720/25-239

23. Gong, Jingzhi, and Tao Chen. "Does configuration encoding matter in learning software performance?: an empirical study on encoding schemes." MSR '22: Proceedings of the 19th International Conference on Mining Software Repositories, Pages 482-494, Oct 2022. https://doi.org/10.1145/3524842.3528431
24. Joseph, V. Roshan. "Optimal ratio for data splitting." *Statistical Analysis and Data Mining*, 15(4), 531-538, 7 Feb 2022. arxiv.org/pdf/2202.03326.

https://doi.org/10.59720/25-239

APPENDIX

GitHub repository with the code generated for this project github.com/adityaramanathan/energy.

Feature	Median	Standard Deviation	Min	Max
Total energy (in 1,000 BTU units)	70,737.35	50,189.92	1,182.22	506,714.32
House size in sq. ft.	1500	883.22	240	12,100
Number of rooms	6	2.37	1	15
Household member count	2	1.37	1	7
Cost of electricity	1,200.9	813.85	-150.5	10,574.78
Cost of natural gas	433.41	481.95	0	7,470.63
Number of bedrooms	3	1.1	0	6
Number of full bathrooms	2	0.76	0	4
Number of half bathrooms	0	0.52	0	2
Number of other rooms	3	1.62	1	9
Number of weekdays spent at home	5	1.6	0	5
Number of refrigerators	1	0.7	0	9
Number of microwaves	1	0.3	0	3
Number of electronic devices	8	4.24	0	42
Number of ceiling fans	2	2.02	0	15
Number of lights 1-4 hours a day	4	6.08	0	90
Number of lights 4-8 hours a day	2	4.62	0	84
Number of lights > 8 hours a day	0	3.75	0	99
Heating degree days	4,221	2,260.55	0	14,111
Cooling degree days	1,256	1,190.09	0	5,414

Table A1: Summary Statistics for Household Features. This table provides a statistical overview of the features from the RECS dataset of 18,496 households used in the machine learning models. For each feature, the median, standard deviation, minimum, and maximum values are displayed.