# Improving Building Energy Efficiency through Data Analysis

DiAndra Phillip
City College of New York
New York, USA
dphilli004@citymail.cuny.edu

Jin Chen
Nearabl Inc.
New York, USA
jin@nearabl.com

Fani Maksakuli
Nearabl Inc.
New York, USA
fani@nearabl.com

Arber Ruci
Nearabl Inc.
New York, USA
arber@nearabl.com

E'edresha Sturdivant
Nearabl Inc.
New York, USA
eedresha@nearabl.com

Zhigang Zhu
City College of New York
New York, USA
zzhu@ccny.cuny.edu

## Abstract

For many lawmakers, energy-efficient buildings have been the main focus in large cities across the United States. Buildings consume the largest amount of energy and produce the highest amounts of greenhouse emissions. This is especially true for New York City (NYC)'s public and private buildings, which alone emit more than two-thirds of the city's total greenhouse emissions. Therefore, improvements in building energy efficiency have become an essential target to reduce the amount of greenhouse gas emissions and fossil fuel consumption. NYC's buildings' historical energy consumption data was used in machine learning models to determine their ENERGY STAR scores for time series analysis and future prediction. Machine learning models were used to predict future energy use and answer the question of how to incorporate machine learning for effective decision-making to optimize energy usage within the largest buildings in a city. The results show that grouping buildings by property type, rather than by location, provides better predictions for ENERGY STAR scores.

***CCS Concepts:*** • **Applied computing** → **Forecasting**; • **Computing methodologies** → *Feature selection*; • **Networks** → *Network performance evaluation.*

***Keywords:*** Building Energy Efficiency, Data Analysis, Machine Learning


*e-Energy '23 Companion, June 20–23, 2023, Orlando, FL, USA*

ACM ISBN 979-8-4007-0227-3/23/06…$15.00
https://doi.org/10.1145/3599733.3600244

**ACM Reference Format:**
DiAndra Phillip, Jin Chen, Fani Maksakuli, Arber Ruci, E'edresha Sturdivant, and Zhigang Zhu. 2023. Improving Building Energy Efficiency through Data Analysis. In *The 14th ACM International Conference on Future Energy Systems (e-Energy '23 Companion), June 20–23, 2023, Orlando, FL, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3599733.3600244

## 1 Introduction

Most of the city's greenhouse gas emissions come from buildings, with higher temperatures, more frequent and intense rainfall, and rising seas chipping away at New York's coastal edges. The City Council passed Local Law 97 in 2019 aimed at reducing the greenhouse gas (GHG) emissions that cause climate change. There are about 12 months until the deadline for building owners to meet the first GHG limits, and there is a threat of fines that could climb to millions of dollars a year for buildings that do not comply. Some buildings are owned by large corporations, while others are run by families and mom-and-pop small investors. Real estate companies with large portfolios are on track to avoid penalties, but mom-and-pop companies that own older buildings are still trying to figure out what they need to do and how they'll pay for these projects [3].

Meeting greenhouse gas emission limits is easier for newly constructed buildings, but for older buildings, especially residential buildings, compliance is difficult. Many buildings still have oil-burning furnaces, and obtaining the funds for a complete renovation is challenging [12]. Building owners need help identifying building processes to reduce emissions. Some are just starting out in the exploratory phases, while others have projects in the works. However, confusion about how the law works and uncertainty about the next steps, including how to pay for upgrades, has been the experience so far for many property owners and co-op boards. Most building owners to whom the law applies must comply by 2024, and citywide, about 70% of buildings are expected to be ready as they are, according to a November report from the New York City Comptroller [15]. However, for many of those same buildings, starting in 2030 when emissions limits decrease, only 30% of the buildings will already be in compliance.

The ENERGY STAR rating [14] is NYC's measure of a building's energy efficiency. It is determined using the United States Environmental Protection Agency (EPA) online benchmarking tool, ENERGY STAR Portfolio Manager [13], which compares a building's energy performance to similar buildings in similar climates.

If a building's energy consumption can be reduced through data analysis, it can help lawmakers develop transitional initiatives to make the city carbon neutral. This analysis provides insights into how implementing or replacing equipment can minimize energy waste and emissions, ultimately improving the energy score. Considering the high cost of adding new systems to buildings, understanding how to maximize energy efficiency while keeping costs low offers an economic advantage. These changes would enhance the quality of life for all New Yorkers and address the challenges posed by climate change. Potential measures may involve identifying green and sustainable alternatives for outdated equipment in various buildings. The objective of this study is to present a new algorithm with significant features applicable to any NYC building, aiming to reduce greenhouse gas emissions.

The primary contribution of this work includes the following:

1. Identified the most significant factors impacting building energy efficiency based on NYC Energy and Water Data Disclosure.
2. Utilized classification and regression models to predict the ENERGY STAR score of a building based on specified parameters, in order to understand the impact of each factor on determining the ENERGY STAR score.
3. Employed time series analysis on building data from 2014 to 2020 and predicted its value in 2021 using Long Short-Term Memory (LSTM) and Facebook Prophet models.

The paper is organized as follows. Section 2 discusses the state-of-the-art models. Section 3 provides an overview of the dataset. Next, Section 4 describes the methods applied in the evaluation, and Section 5 compares and discusses the results. Finally, Section 6 concludes the paper and provides potential further research.

## 2 Related Work

The integration of artificial intelligence (AI) in big cities aids in shaping energy consumption patterns and planning for the future. These cities consume large amounts of energy and generate GHG emissions, which poses a persistent challenge to address. The reduction of emissions can help improve urban air quality and mitigate the frequency of disastrous weather events. Various studies on building energy efficiency utilize artificial neural networks and cluster-based methods to develop predictive models. Research based on the Commercial Buildings Energy Consumption Survey [18] suggests that grouping data into smaller samples yields better predictions. When buildings are grouped by property type, the artificial neural network model exhibits lower mean squared error compared to multiple linear regression approaches for each specific property type.

Artificial neural networks (ANN) have emerged as the primary method for predicting energy consumption. Mena et al. [4] utilized ANN to estimate Spain's CIESOL bioclimatic building electricity demand. The study demonstrated that outdoor temperature and solar radiation significantly impact electricity consumption. Similarly, Yalcintas et al. [18] employed ANN to predict electricity consumption per square meter (EUI) based on energy benchmarking data from the U.S. Commercial Buildings Energy Consumption Survey, considering climate variations. The model incorporated physical properties and occupancy information for office-type buildings. To enhance accuracy, categorical variables, such as lighting and cooling percentages, were converted from nominal to categorical before developing the ANN model. The predictive accuracy of EUI was compared with multiple linear regression methods, showcasing a significant advantage over simple linear regression.

Clustering algorithms have been employed to provide detailed insights into building energy usage. Petcharat et al. [9] utilized three different methods to estimate potential energy savings in lighting systems in Thailand. The first two methods compared the light power density (LPD) and average LPD of each building to the target, while the third method employed the Expectation Maximization algorithm for clustering. The research demonstrates that cluster-based analysis (with an error range of 0-11%) outperforms the averaging method, which consistently underestimated potential savings (with an error range of 1-100%) due to its susceptibility to outliers. Gao et al. [2] also demonstrate that employing clustering algorithms with energy benchmarking data yields higher accuracy compared to relying solely on the ENERGY STAR score. The clustering approach takes into account all relevant building features that impact energy consumption, whereas the ENERGY STAR approach merely categorizes buildings based on use types without considering the influence of other building features. The proposed methodology clusters buildings based on their total energy performance and climate differences.

The aforementioned research offers valuable insights into predictive models for building energy efficiency. However, these studies do not address the challenges faced by building owners in implementing specific changes to improve energy efficiency effectively. The use of the ENERGY STAR rating alone is not sufficient. Scofield et al.[11] analyzed the ENERGY STAR models for typical building types and found that the scores produced by their models have uncertainties of

**Table 1.** ENERGY STAR score by letter grade.

| ENERGY STAR Score |
|---|
| A – score is equal to or greater than 85<br>B – score is equal to or greater than 70 but less than 85<br>C – score is equal to or greater than 55 but less than 70<br>D – score is less than 55<br>F – Benchmarking information not submitted |

±35 points. The study concludes that there is no justification for quantitative claims of energy savings or reduction in GHG emissions based solely on ENERGY STAR scores. This paper aims to fill that gap by analyzing NYC energy and water benchmarking data and developing an algorithm to understand the impact of different building features on energy consumption and greenhouse gas emissions.

# 3 Dataset

The NYC Mayor's Office of Climate and Environmental Justice conducts an annual data collection process through the EPA ENERGY STAR Portfolio Manager, gathering over 29,000 building metrics related to water and energy consumption. This data encompasses privately owned buildings exceeding 25,000 sq ft and City-owned buildings exceeding 10,000 sq ft [7]. The primary purpose of this data collection process is to enable building owners to measure and compare their energy and water consumption with similar structures in the city, promoting transparency in building energy and water usage [8]. The collected data includes various building metrics such as water consumption, electric consumption, gas consumption, and greenhouse gas (GHG) emissions. Additionally, each building is assigned a ENERGY STAR score based on its energy use compared to the best and worst performing buildings, as well as those in between. This score is calculated by comparing the estimated energy values with the actual energy data provided, allowing for benchmarking against peer buildings within the same property type group [14].

The present research utilized data obtained from the Energy and Water Data Disclosure, sourced from NYC Open Data [6], spanning the years 2014 to 2021. The original dataset included a substantial number of buildings with diverse property types. However, for the purpose of our study, we focused on investigating the relationship between property types and locations. Consequently, our analytical sample comprised 3777 residential buildings in the Bronx, 1491 office buildings in Manhattan, and 513 educational buildings in Brooklyn. The selection of these buildings was guided by specific criteria to ensure variations in location and primary property type, enabling us to compare significant features between the groups.

# 4 Method

The objective of this research is to identify the most significant factors that impact building energy efficiency and to forecast future energy wastage by considering the existing equipment and potential effects of equipment replacement on energy wastage and emissions. The research methodology encompasses several stages. Firstly, classification and regression models will be utilized to predict the ENERGY STAR score of a building, taking into account specified parameters. This analysis aims to comprehend the impact of each factor on determining the ENERGY STAR score. Secondly, time series analysis will be employed using building data from 2014 to 2020, enabling the prediction of ENERGY STAR score values for the year 2021. Lastly, the influence of property type and location on ENERGY STAR score evaluation will be investigated. This step aims to examine how these factors contribute to the overall energy performance of buildings.

## 4.1 Feature Extraction

This research utilized the 2014 to 2021 energy and water data, which initially consisted over 29,000 buildings and 250 building attributes for each year. The primary objective was to comprehend the building energy usage by examining the characteristics of these buildings. The data was initially provided in JSON format, with all variables represented as strings. Consequently, several attributes in the dataset are technically numeric, but they were initially represented as strings. Therefore, it was necessary to convert these attributes from string to numeric type in order to ensure their appropriate treatment and analysis. Additionally, the dataset contained instances of "Not Available" and "Insufficient access" in many cells, which were replaced with "NAN." As a result, the columns containing only numeric values were successfully converted to the appropriate numeric type.

There were several building attributes in the dataset that had more than 75% missing data. To maintain the integrity of the sample, those columns were dropped, as filling them with a descriptive statistic would not accurately represent the data. On the other hand, for the remaining columns that had less than 10% missing data, the missing cells were filled with the column mean. Furthermore, a new indicator value was added to identify the rows where imputation was performed.

To address the presence of potential outliers, the z-score function was applied to filter out buildings' attributes that deviated more than three standard deviations from the mean. Following this cleaning process, approximately 22,000 buildings and 89 building attributes remained in the dataset. Additionally, the data were scaled using standardization and normalization techniques to account for variations in unit measurements across different building consumption metrics, which span a range from tens to millions. The performance of these scaled values will be compared to that of the raw data to obtain the most optimal results.

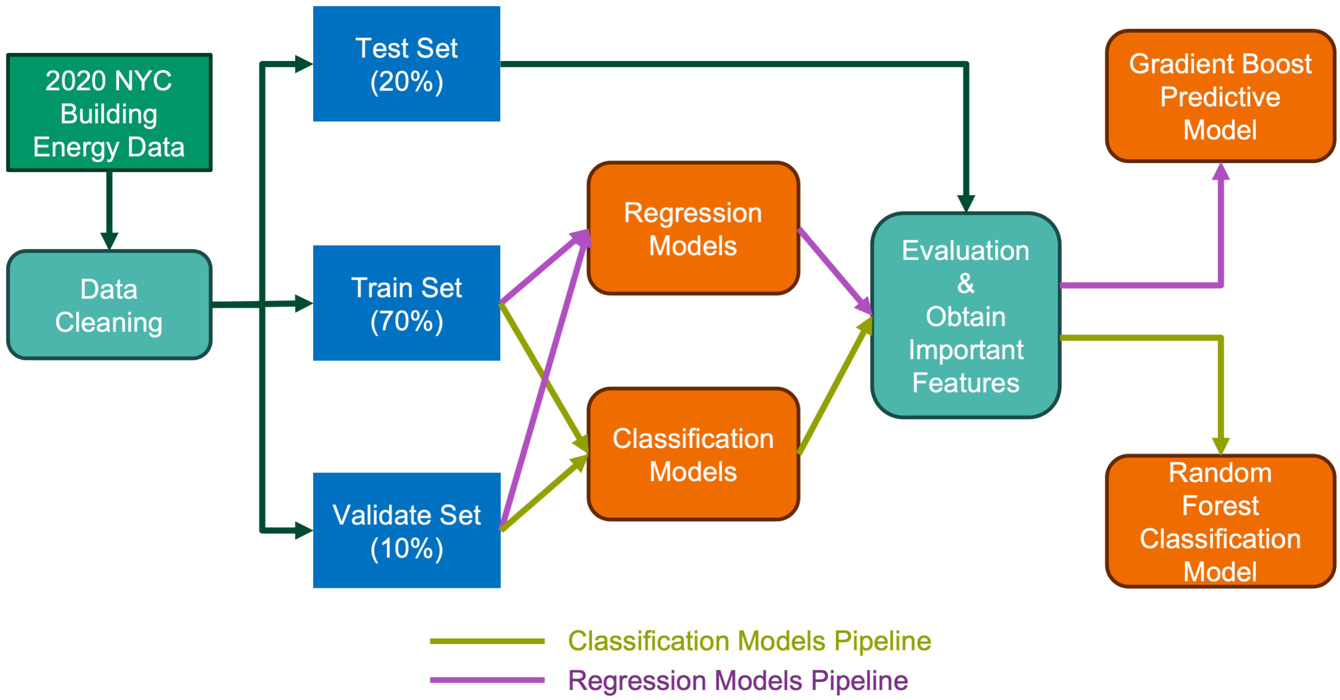


**Figure 1.** Flowchart: classification and regression model training and evaluation for singer year data.

### 4.2 Building ENERGY STAR Score Estimation

After completing the preprocessing phase, three distinct sample sets were carefully chosen to construct machine learning models. These sets were specifically comprised of residential buildings in the Bronx (3777 buildings), office buildings in Manhattan (1491 buildings), and educational buildings in Brooklyn (513 buildings), as elaborated in Section 3. To ensure an effective evaluation process, the dataset was split into randomized train-test-validation subsets. The training set was allocated 70% of the data and utilized for model training. A separate validation set, encompassing 10% of the data, was employed for fine-tuning the network architecture and optimizing the model's parameters to achieve optimal performance. Lastly, the remaining 20% of the data constituted the test set, which was utilized to comprehensively assess the performance of the trained models.

The data was transformed into a classification problem by replacing the ENERGY STAR score with its corresponding letter grade rating, as shown in Table 1. The higher the rating, the better the energy efficiency performance [7]. The models used for classification included Decision Trees, Extreme Gradient Boosting, Support Vector Machine (SVM), and Random Forest.

The regression methods employed to identify significant features for energy consumption in public buildings included AdaBoost, Decision Trees, Extreme Gradient Boosting, SVM, and Random Forest. These methods were selected for their capacity to capture nonlinearity and effectively learn from historical data. The objective was to extract important predictors that could serve as inputs for time series forecasting. The design utilized for each model is illustrated in Figure 1.

During the hyperparameter tuning process for each model, various parameters were adjusted to achieve improved results. For the AdaBoost and Extreme Gradient Boosting models, the main parameters that were modified included the number of estimators, subsample, maximum depth, and the learning rate. In the case of the Decision Tree and Random Forest models, the parameters that were adjusted were the number of estimators, maximum depth, maximum leaf nodes, and minimum samples per leaf. Lastly, for the SVM model, the regularization parameters (C and gamma) as well as the choice of kernel function were fine-tuned.

### 4.3 Time Series Analysis

The second objective of this research was to predict building ENERGY STAR scores using time series data. For the analysis, the energy and water dataset spanning from 2014 to 2021 was divided into train, test, and validation sets following the same ratio as described in Section 4.2.

Three baseline models (KNN, decision trees, and linear regression) were employed along with deep learning models (LSTM and Prophet) for the analysis.

The baseline models utilized the Skforecast Python library, which leverages scikit-learn regressors as multi-step forecasters [10]. These models served as a benchmark for comparison against the more advanced deep learning models.

LSTM was chosen for its ability to capture important information from a sequence while disregarding less significant details, while Prophet was selected for its capability to analyze trend and seasonality [16]. Utilizing the relevant features identified from the regression analysis, time series models were developed to predict a building's energy score based on past years' data. Prior to constructing the models as depicted in Figure 2, a data imbalance issue was addressed where certain boroughs contained a larger number of buildings. To ensure consistency, a general matrix was created that maintained the same set of buildings for comparing the general model with the models specific to boroughs and property types.

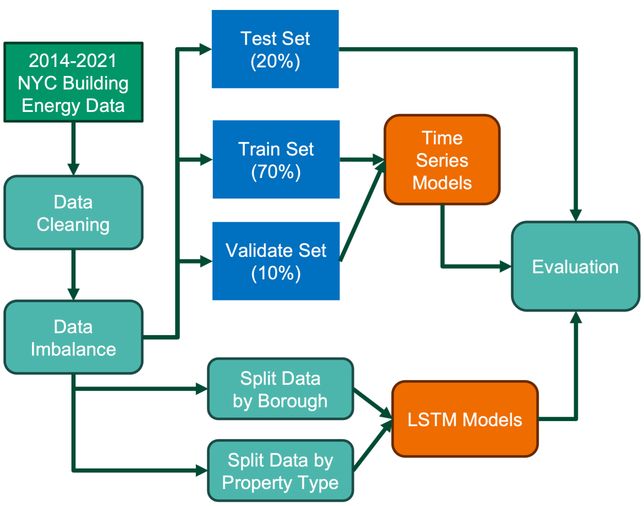


**Figure 2.** Flowchart for predicting ENERGY STAR score with time series models.

To create the general matrix, the data from 2014 to 2021 underwent the same preprocessing steps as described in Section 4, with the exception of removing outliers and variable reduction since the variables used were extracted features as explained above. Property ids were also checked to ensure that a property was present throughout each year, and properties that were not present were excluded from the analysis. This process resulted in data imbalances, with approximately 10,000 buildings remaining, the majority of which belonged to three different property types: multifamily, educational, and offices, out of a total of nine property types. With the remaining data, bootstrap sampling was applied to create equally-sized subsets of data based on a building's location and property type, ensuring a balanced dataset.

## 5 Experiments

### 5.1 Building Attributes and ENERGY STAR Score Correlation

The permutation feature importance is a model inspection technique that proves useful for non-linear estimators. In this process, the relationship between the feature and the target is disrupted, whereby a single feature value is randomly shuffled. The decrease in the model's score indicates the extent to which the model relies on the feature [1]. Feature importance was calculated on the validation set to highlight the features that contribute the most to the model. This process aided in reducing the remaining 89 features necessary for ENERGY STAR score prediction. Although the weights of these features differed across datasets (Bronx residential buildings, Manhattan office buildings, and Brooklyn educational buildings), all three shared energy usage and site electricity as the top 2 predictors. Longitude and latitude are important predictors since building density within an area can impact energy usage [5]. This affects the amount of solar energy a building can receive. After comparing the 89 features within each dataset, the 12 features were identified as being significant for future prediction. Table 3 displays the top 5 important features of the building type and borough. The remaining 7 important features include occupancy, number of buildings, natural gas use (kBTU), eGrid output emissions rate((kgCO2e/MBtu)), offsite green power (kWh), and the previous year's score.

**Table 2.** Mean absolute error (MAE) of building ENERGY STAR score estimation by regression models: original data, normalized data, and standardized data comparison.

| Model | Original | Normalized | Standardized |
|---|---|---|---|
| AdaBoost | 11.43 | 16.66 | 12.13 |
| Decision Trees | 8.59 | 12.04 | 8.63 |
| Gradient Boosting | 6.05 | 8.51 | 6.05 |
| SVM | 24.9 | 22.58 | 19.73 |
| Random Forest | 5.59 | 8.67 | 5.58 |

To estimate the ENERGY STAR score of the buildings using the available data, we employed various regression models (i.e., AdaBoost, Decision Trees, Extreme Gradient Boosting, SVM, and Random Forest) on three sample sets: residential buildings in the Bronx, office buildings in Manhattan, and educational buildings in Brooklyn, obtained from the Energy and Water Data Disclosure for 2020.

Table 2 displays the average testing error between the predicted ENERGY STAR score and the actual value. Among the models, gradient boosting and random forest with the lowest errors. These two models were further utilized for parameter tuning. After hyperparameter tuning, the random forest model yielded the best results, with a mean absolute error of approximately 5.6. Consequently, the random forest model was selected for comparing the significant features across the individual sample datasets.

**Table 3.** Top 5 significant features and their corresponding importance scores extracted from regression models.

| Feature | Importance Score |
|---|---|
| Electricity Use | 0.52 |
| Gross Floor Area | 0.10 |
| Latitude | 0.09 |
| Year Built | 0.07 |
| Longitude | 0.07 |

**Table 4.** Distribution of buildings by boroughs and property types in time series model dataset.

| Borough | Residential | Office | Educational |
|---|---|---|---|
| Bronx | 50 | 30 | 50 |
| Brooklyn | 50 | 30 | 50 |
| Manhattan | 50 | 30 | 50 |
| Queens | 50 | 30 | 50 |
| Staten Island | 50 | 30 | 50 |

### 5.2 Time Series Analysis

Prior to conducting various model tests, several data imbalances were observed where the majority of the buildings were of office, residential, or educational property type. Specifically, there was a notable disparity in the number of residential and office buildings compared to other building types. Similarly, when comparing buildings based on location, there was a marked overrepresentation of buildings located in Manhattan and Brooklyn in comparison to those in Staten Island or Queens. To address this concern, a bootstrapping technique discussed in Section 4.3 was applied on the original dataset, narrowing the focus to buildings categorized as residential, office, and educational, as illustrated in Table 4. Due to insufficient data in multiple locations for the other property types, sample subsets were created to achieve a balanced dataset

In conducting our time series analysis, we utilized data spanning from 2014 to 2020 and generated forecasts for 2021. The input data consisted of the variables from feature extraction along with the computed difference in the ENERGY STAR score of the last 2 years and the output is the predicted ENERGY STAR score for 2021. Our study incorporated five distinct time series models discussed in Section 4.3, consisting of three baseline models - KNN, decision trees, and linear regression - as well as two deep learning models, LSTM and Prophet. The results are outlined in Table 5. It is worth noting that while both the linear regression and LSTM models exhibited comparable performance, the linear regression model's predictions for unseen data were based on the mean value of all buildings, rendering the results inconsequential.

**Table 5.** Performance comparison of time series models for forecasting building ENERGY STAR scores. (MAE: mean absolute error, RMSE: root mean square error)

| Models | MAE | RMSE |
|---|---|---|
| KNN | 27.10 | 35.11 |
| Decision Trees | 26.48 | 31.56 |
| Linear Regression | 21.22 | 25.55 |
| LSTM | 24.15 | 29.88 |
| Prophet | 28.27 | 35.31 |

The prediction of KNN and decision trees produced a repetitive continuous pattern as shown in Figure 3. This may be due to the model assuming seasonality in the data since the ENERGY STAR score appears to repeat an increased score to a decreased score between the buildings. The prediction of prophet and LSTM is more adaptable and able to predict the buildings' large range in energy score. Although prophet is more adaptable even after parameter tuning the predicted ENERGY STAR score was not better than LSTM predictions. Consequently, LSTM was deemed the optimal model for the purposes of this analysis.

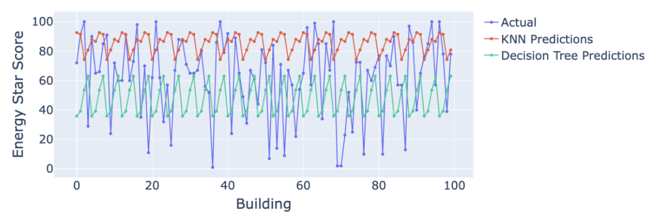


**Figure 3.** The predicted ENERGY STAR score of the baseline models compared to actual values.

#### 5.2.1 Location-based Analysis.

Rather than conducting an analysis of buildings across all five boroughs of New York City, we trained individual LSTM model specifically for each borough. As a result, all five borough-based LSTM models have better performance in comparison to the general LSTM model. The general matrix predictions with all the buildings together had a mean absolute error of 24.15. The results in Table 6 show a decrease of approximately 6 or more in the mean absolute errors (MAE) between the models by location.

Although there is a significant decrease when considering the letter grade the error range is still too high for an ENERGY STAR score. Taking a closer look the high errors may be due to errors in the dataset. It is possible the raw inputs may be off by a number causing an extreme difference between the actual and predicted. The model predicts accurately for several buildings. The high error is due to the difference in less than 10% of buildings in the models. With more data in the models, the prediction can be better.

**Table 6.** Performance comparison of borough-based LSTM models for forecasting building ENERGY STAR scores. (MAE: mean absolute error, RMSE: root mean square error)

| Borough | MAE | RMSE |
|---|---|---|
| Bronx | 12.28 | 17.27 |
| Brooklyn | 16.03 | 21.40 |
| Manhattan | 16.97 | 21.48 |
| Queens | 18.96 | 24.80 |
| Staten Island | 13.29 | 17.45 |

**Table 7.** Performance comparison of property type-based LSTM models for forecasting building ENERGY STAR scores. (MAE: mean absolute error, RMSE: root mean square error)

| Property Type | MAE | RMSE |
|---|---|---|
| Residential | 13.66 | 18.22 |
| Office | 10.54 | 15.30 |
| Educational | 16.02 | 23.69 |

5.2.2 **Property Type Analysis.** The general matrix was also split into property types to evaluate if the LSTM model performs better. Overall the models performed better than the general matrix. Looking at Table 7 the office model has the lowest error; accurately predicting most of the buildings' energy scores. The general matrix predicting the building without grouping by property type has a higher mean absolute error of at least 8.

Similar to the borough-based models, the property type-based models demonstrate accurate predictions for some buildings, while others exhibit significant deviations from the actual values. Despite a few buildings with substantial errors, the property type models outperform other models in terms of overall performance.

## 6 Conclusion

In conclusion, our study employed feature extraction to identify predictive variables for estimating the ENERGY STAR score. We also developed models for score estimation and conducted time series analysis using multiple techniques. Through the feature extraction process, we successfully identified the most useful variables for predicting the ENERGY STAR score, resulting in a reduction of the original set of over 250 attributes for each building to a final selection of 12 attributes. By understanding how those 12 attributes can influence energy efficiency can have a positive impact on a building owner's decision-making process regarding future improvements in energy usage.

When estimating the ENERGY STAR score, the comparison of several regression models consisting of AdaBoost, Decision Trees, Extreme Gradient Boosting, SVM, and Random Forest, along with 2 different scaling techniques (standardization and normalization), was performed. This analysis revealed that the random forest model is better and more accurately estimated the ENERGY STAR score, with the lowest error.

In the time series analysis, KNN and Decision Trees models had a repetitive pattern in their predictions, while Linear Regression model predicted the average value of all the buildings. Among Prophet and LSTM models, LSTM demonstrated the best performance by accurately predicting the ENERGY STAR scores of several buildings. The data also showed higher predicted scores when the actual values were lower, which could be attributed to potential errors in the dataset, as it is submitted by building owners. The results highlight that separating the data by property type yields the highest performance in predicting their ENERGY STAR score, suggested there might be different standard for energy efficiency for different property type.

### 6.1 Future Work

For future research, obtaining more precise and comprehensive historical energy consumption data from various property types will be crucial in achieving more accurate results. The integration of third-party data, which includes information on HVAC equipment and boiler types, would prove valuable in enabling better predictions and identifying areas where building owners can focus their efforts. Additionally, including additional attributes such as the number of floors, total units in multifamily buildings, building construction, and environmental information (such as external climate conditions) as additional inputs would enhance the models' predictive power. Weather conditions, in particular, play a significant role in determining and predicting building energy usage. Some studies have proposed simplifying weather conditions in building energy calculations by utilizing average monthly temperatures, as attempted by White and Reichmuth [17]. Incorporating these factors can significantly improve the models' predictions and assist building owners in more effectively complying with LL97.

Based on the findings and methodology presented, future studies could build upon this research by employing similar machine learning models to predict ENERGY STAR scores for buildings in various big cities, grouped by property type. This approach would enable the identification and targeting of low-performing buildings, as well as the estimation of future ENERGY STAR scores. Energy disclosure data, coupled with external sources like data from utility providers, serve as crucial resources for long-term greenhouse gas reduction and energy efficiency initiatives in big cities.

## Acknowledgments

The work is supported by the US National Science Foundation (#2131186, #1737533, #1827505, #1644664, and #2048498), ODNI Intelligence Community Center for Academic Excellence (IC CAE) at Rutgers (#HHM402-19-1-0003 and #HHM402-18-1-0007) and the US Air Force Office for Scientific Research (#FA9550-21-1-0082).

## References

[1] Scikit-Learn Developers. 2023. 4.2. permutation feature importance. https://scikit-learn.org/stable/modules/permutation_importance.html

[2] Xuefeng Gao and Ali Malkawi. 2014. A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm. *Energy and Buildings* 84 (2014), 607–616. https://doi.org/10.1016/j.enbuild.2014.08.030

[3] Jane Margolies. 2022. *New York developers rush to reduce emissions as hefty fines loom*. https://www.nytimes.com/2022/08/16/business/new-york-real-estate-climate-change.html

[4] R. Mena, F. Rodríguez, M. Castilla, and M.R. Arahal. 2014. A prediction model based on neural networks for the energy consumption of a bioclimatic building. *Energy and Buildings* 82 (2014), 142–155. https://doi.org/10.1016/j.enbuild.2014.06.052

[5] Nariman Mostafavi, Mehdi Pourpeikari Heris, Fernanda Gándara, and Simi Hoque. 2021. The Relationship between Urban Density and Building Energy Consumption. *Buildings* 11, 10 (2021). https://doi.org/10.3390/buildings11100455

[6] Mayor's Office of Climate and Sustainability. 2021. Energy and water data disclosure for Local Law 84 2021 (data for calendar year 2020): NYC open data. https://data.cityofnewyork.us/Environment/Energy-and-Water-Data-Disclosure-for-Local-Law-84-/usc3-8zwd

[7] Mayor's Office of Climate and Sustainability. 2023. Energy grades. https://www.nyc.gov/site/buildings/property-or-business-owner/energy-grades.page

[8] City of New York. 2023. Benchmarking and energy efficiency rating. https://www.nyc.gov/site/buildings/codes/benchmarking.page

[9] Siriwarin Petcharat, Supachart Chungpaibulpatana, and Pattana Rakkwamsuk. 2012. Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings. *Energy and Buildings* 52 (2012), 145–152. https://doi.org/10.1016/j.enbuild.2012.06.006

[10] Joaquín Amat Rodrigo and Javier Escobar Ortiz. 2021. Skforecast: Time series forecasting with Python and Scikit-Learn. https://www.cienciadedatos.net/documentos/py27-time-series-forecasting-python-scikitlearn.html

[11] John Henry Scofield. 2016. . John H. Scofield.

[12] Somini Sengupta. 2022. An oily challenge: Evict stinky old furnaces in favor of heat pumps. https://www.nytimes.com/2022/09/14/climate/oil-gasfurnace-heat-pump.html

[13] ENERGY STAR. 2023. Benchmark Your Building Using ENERGY STAR® Portfolio Manager®. https://www.energystar.gov/buildings/benchmark

[14] Energy Star. 2023. How the 1-100 energy star score is calculated. https://www.energystar.gov/buildings/benchmark/understand_metrics/how_score_calculated

[15] Jamie Statter, Annie Levers, Louise Yeung, and Celeste Hornbach. 2022. Strong Implementation of Local Law 97, NYC's Green New Deal for Buildings. https://comptroller.nyc.gov/reports/cap-the-credits/

[16] Sean J Taylor and Benjamin Letham. 2012. Forecasting at scale. *PeerJ Preprints 5:e3190v2* 52 (2012), 145–152. https://doi.org/peerj.preprints.3190v2

[17] J A White and H Reichmuth. 1996. Simplified method for predicting building energy consumption using average monthly temperatures. (12 1996). https://www.osti.gov/biblio/474419

[18] Melek Yalcintas and U. Aytun Ozturk. 2007. An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database. *International Journal of Energy Research* 31, 4 (2007), 412–421. https://doi.org/10.1002/er.1232 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/er.1232