universität koblenz
weiter:denken
Faculty 4: Computer Science

WeST
Web Science and Technologies
Institute for Web Science
and Technologies

Explainable Artificial Intelligence for Site Energy Usage Intensity Prediction

Master's Thesis

in partial fulfillment of the requirements for
the degree of Master of Science (M.Sc.)
in Informatik

submitted by
Aman Singla

First supervisor:	Prof. Dr. Frank Hopfgartner Institute for Web Science and Technologies
Second supervisor:	Dr.-Ing. Stefania Zourlidou Institute for Web Science and Technologies

Koblenz, September 2024

Statement

I hereby certify that this thesis has been composed by me and is based on my own work, that I did not use any further resources than specified – in particular no references unmentioned in the reference section – and that I did not submit this thesis to another examination before. The paper submission is identical to the submitted electronic version.

	Yes	No
I agree to have this thesis published in the library.	☑	☐
I agree to have this thesis published on the Web.	☑	☐
The thesis text is available under a Creative Commons License (CC BY-SA 4.0).	☑	☐
The source code is available under a GNU General Public License (GPLv3).	☑	☐
The collected data is available under a Creative Commons License (CC BY-SA 4.0).	☑	☐

Koblenz, September 20, 2024 ..... Aman

(Place, Date) (Signature)

Note

- If you would like us to contact you for the graduation ceremony,
  please provide your personal E-mail address: amansingla444@gmail.com
- If you would like us to send you an invite to join the WeST Alumni
  and Members group on LinkedIn, please provide your LinkedIn ID : linkedin.com/in/aman-singla/

Zusammenfassung

Künstliche Intelligenz (KI) bestimmt in zunehmendem Maße, wie unsere täglichen Erfahrungen gestaltet werden. Der weit verbreitete Einsatz von KI-Modellen in verschiedenen Bereichen hat Bedenken hinsichtlich möglicher Verzerrungen in diesen Modellen geweckt und zu einer Forderung nach mehr Transparenz und Interpretierbarkeit geführt. Die Herausforderungen bei der Interpretation komplexer maschineller Lernmodelle haben KI-Forscher und -Praktiker dazu veranlasst, ihren Schwerpunkt auf erklärbare KI (Explainable AI, XAI) zu verlagern, um das Verständnis zu verbessern und das Vertrauen in diese Modelle zu stärken, selbst wenn sie für umfangreiche Anwendungen eingesetzt werden.

Das Hauptziel dieser Arbeit ist es, die verschiedenen Faktoren zu untersuchen, die zum Energieverbrauch von Gebäuden beitragen, und Modelle zu entwickeln, die den Energieverbrauch von Gebäuden vorhersagen und gleichzeitig den Entscheidungsprozess dieser Algorithmen erklären. Der in dieser Arbeit verwendete Datensatz besteht aus Variablen, die sich auf Gebäudeeigenschaften, Klima und Wetterbedingungen in verschiedenen Regionen beziehen. Genaue Vorhersagen des Energieverbrauchs sind wichtig, um politischen Entscheidungsträgern dabei zu helfen, Initiativen zur Gebäudesanierung strategisch auszurichten und dadurch eine optimale Reduzierung der Treibhausgasemissionen zu erreichen.

Die Ergebnisse dieser Studie zeigen, dass der Random-Forest-Algorithmus (RF) im Vergleich zu anderen Algorithmen des maschinellen Lernens die genauesten Vorhersagen liefert. Zu den wichtigsten Treibern des Energieverbrauchs, die durch XAI-Techniken wie SHAP und LIME identifiziert wurden, gehören das Energy-Star-Rating, der Gebäudetyp und die Grundfläche. Diese XAI-Methoden trugen dazu bei, die Interpretierbarkeit der Modelle zu verbessern, so dass sie auch für nicht fachkundige Nutzer wie Gebäudemanager und politische Entscheidungsträger leichter zugänglich sind. Durch den Einsatz von maschinellem Lernen und XAI bietet diese Forschung einen transparenten und umsetzbaren Rahmen für die Optimierung der Energieeffizienz von Gebäuden und die Unterstützung eines nachhaltigen Energiemanagements.

Abstract

Artificial Intelligence (AI) is increasingly guiding how our daily experiences are shaped. The widespread use of AI models across various domains has raised concerns about potential biases in these models and has led to a demand for greater transparency and interpretability. The challenges of interpreting complex machine learning models have prompted AI researchers and practitioners to shift their focus towards explainable AI (XAI), seeking to enhance understanding and build trust in these models even when applied to large-scale applications.

The primary goal of this thesis is to examine the diverse factors contributing to building energy usage and develop models that predict building energy consumption while explaining the decision-making process of these algorithms. The dataset used in this research consists of variables related to building characteristics, climate, and weather conditions across different regions. Accurate predictions of energy consumption are important for helping policymakers to strategically target building renovation initiatives and thereby achieve optimal reductions in greenhouse gas emissions.

The findings of this study demonstrate that the Random Forest (RF) algorithm provided the most accurate predictions in comparison with other boosting machine learning algorithms. Key drivers of energy consumption identified through XAI techniques such as SHAP and LIME include energy star rating, facility type, and floor area. These XAI methods helped enhance the interpretability of the models, making them more accessible for non-expert users, such as building managers and policymakers. By leveraging machine learning and XAI, this research provides a transparent and actionable framework for optimizing building energy efficiency and supporting sustainable energy management.

Acknowledgement

I would like to express my deepest gratitude to the Institute of Web Science and Technologies and its esteemed members. I am thankful to my first supervisor, Prof. Dr. Frank Hopfgartner, and my second supervisor, Dr. Ing. Stefania Zourlidou, for their invaluable guidance and support throughout the development of this thesis.

I am incredibly grateful to Dr. Ing. Stefania Zourlidou for her dedicated involvement at every step of the process - from helping me brainstorm and narrow down my ideas into a focused topic, assisting in the formulation of my thesis proposal, and solving my queries in regular biweekly meetings. Her encouragement, timely feedback, and patience in addressing my doubts helped me remain focused and motivated throughout this journey.

I would also like to thank my family and friends for their unwavering encouragement and support during the challenging phases of my research.

Finally, my sincere thanks go to the University of Koblenz for providing the academic resources and an enriching environment that enabled me to successfully complete this research.

Contents

List of Figures

List of Tables

1 Introduction

Climate change is a complex and pressing issue that requires both mitigation and adaptation. Mitigation involves reducing greenhouse gas emissions, while adaptation involves preparing for the unavoidable consequences of climate change [44]. This requires changes in electricity and heating systems, ways of transport, industries and buildings to address climate change. According to the International Energy Agency (IEA) [31], buildings, in particular, are responsible for a significant portion of global energy-related and process-related CO2 emissions while buildings from construction to demolition were responsible for 37 percent of global energy-related and process-related CO2 emissions in 2020. However, it is possible to reduce the energy consumption of buildings by implementing retrofitted and modern sustainable neighbourhoods [45]. For instance, retrofitted buildings can reduce heating and cooling energy requirements by 50-90 percent [44]. Not only do these energy-efficient measures reduce costs, but they also improve indoor air quality and maintain the building's functionality.
Accurate and interpretable prediction of building energy consumption holds immense potential for optimizing energy efficiency, reducing costs, and contributing to sustainable development. Traditional prediction models often lack in interpretability, hindering stakeholder understanding and trust in their recommendations. Explainable Artificial Intelligence (XAI) techniques offer a promising solution by unveiling the "why" behind predictions, fostering informed decision-making and targeted interventions. To understand how to make buildings more energy efficient, WiDS Datathon dataset [33] created by Climate Change AI and Lawrence Berkeley National Laboratory is used in this research. This dataset includes information about building characteristics, climate, and weather. This data is used to create an AI model that can predict how much energy (Heating and Electrical) a building will consume. The AI model used here would be explained using various Explainable AI techniques. Explainability in machine learning refers to the process of explaining to a human why and how a machine learning model made a decision [1]. It is the process of analyzing machine learning model decisions and results to understand the reasoning behind the system's decision [11]. Model explainability makes the algorithm's decision-making process transparent to humans [1]. This is particularly important with 'black box' machine learning models, which develop and learn directly from the data without human supervision or guidance [46]. Many machine learning models, despite achieving high levels of precision, are not easily understandable for how a recommendation is made [21]. In case of deep learning models this is especially required to be addressed. As humans, it should must be fully understand how decisions are being made so that the decisions of AI systems can be

trusted.

1.1 Motivation

The necessity of overcoming climate change cannot be overemphasized. At the present moment, this issue stands as one of the most pressing and all-encompassing challenges faced by humanity at large. These impacts can now be felt across the globe, evidenced by intensifying adverse weather events, sea-level rise, and disruptions to ecosystems and human communities alike. The window of meaningful opportunities is closing rapidly, and there has probably never been more need to design final, overarching strategies that reduce gas emissions into the atmosphere.

In this context, the human-built environment emerges as a pivotal domain for consideration. Structures make a significant contribution to global Carbon emissions in their development, operations, and maintenance. They are particularly responsible for about 37 percent of global energy-related and process-related CO2 emissions [49]. This fact underlines how much buildings contribute to the greater climate crisis. A very high proportion of the overall energy use in the world is attributed to the energy used for heating, cooling, lighting, or operational requirements of buildings. As such, any serious endeavor to mitigate climate change has to involve actions that would considerably reduce the energy footprint of buildings.

There is huge potential to save in building energy consumption. The energy savings could be very high if these measures include such actions as retrofitting existing structures with advanced technologies and renewable energy sources, upgrading insulation, and improving heating and cooling systems. Also, emphasis on sustainability in new constructions directly from the design stage can further help reduce overall energy demand. Realizing the full potential of these energy-saving measures would require more than just the application of advanced technologies. This is also a call for the availability of effective tools predicting, managing, and ensuring the optimum utilization of energy in buildings.

This is important for two major reasons: robust prediction of energy use and a need for these predictions to be interpretable and actionable by an eclectic range of stakeholder types.

Explainable Artificial Intelligence (XAI) is one of the most promising solutions and techniques within it's domain are being developed with the aim of rendering AI models' decision-making processes transparent and understandable [25]. By explaining how AI models arrive at their decisions, XAI closes the gap between technically opaque machine learning algorithms and users in the loop [2]. XAI is able to transform those opaque predictions into actionable insights that building managers, policymakers, and other nonexpert users can confidently use to guide decision-making [5]. The use of XAI could go beyond just improving the accuracy of energy consumption predictions. More usability increases with the human interpretability of such predictions. These, in turn, increase usability because when stakeholders can interpret and trust the predictions from AI models, most probably

they undertake informed and effective actions to improve energy efficiency [19]. For instance, this is expected to lead to an increasing uptake of energy-saving measures and consequently contribute to the global effort to mitigate climate change [6]. The power of XAI is enormous and must be applied to overcome the challenges faced in building energy. Focusing on the main drivers of energy use in buildings, the application of XAI techniques aims at developing tools and insights to sustain more sustainable building practices. The goal is to provide stakeholders with the information they need to make decisions that not only reduce energy consumption but also contribute to the broader objective of combating climate change. In doing so, this research seeks to play a role in the global movement towards a more sustainable and resilient future.

1.2 Research Objectives

Based on the need of understanding the ML models this thesis aims to:

- Evaluate the accuracy and interpretability of Machine Learning (ML) models for building energy prediction: This objective assesses the effectiveness of various ML algorithms in predicting overall energy consumption using both heating and electrical data. The study incorporates building characteristics and external factors for comprehensive analysis. The performance of different models will be compared, emphasizing the role of XAI techniques in facilitating interpretation and understanding.
- Identify key drivers of building energy consumption through XAI: Utilizing feature attribution methods within XAI frameworks, this objective delves into identifying the most impactful variables influencing energy use in buildings. The research will explore the influence of diverse factors, including building characteristics (size, age, insulation quality) and external factors (weather, wind speed) on energy consumption, leveraging the insights provided by XAI for deeper understanding.
- Enhance the usability of energy prediction models for non-expert users with XAI: This objective investigates how XAI techniques can be employed to improve the interpretability and usability of energy prediction models for stakeholders without technical expertise. By addressing the user-centric aspect, the research intends to bridge the gap between technical predictions and readily actionable insights, thereby empowering effective decision-making and policy formulation in promoting energy efficiency.

2 Literature Review

As machine learning models become more complex, this has led to an increase in research on Explainable AI (XAI). It is particularly crucial in the field of site energy usage prediction, as it allows users to pinpoint the key factors driving their energy consumption and take proactive steps to improve efficiency.

2.1 Related Work

Sakkas et al's [53] research delves into Explainable Approaches for Forecasting Building Electricity Consumption. The research highlights the growing importance of understanding the factors that influence electricity demand, especially as buildings increasingly incorporate technologies like electric vehicle charging and smart grids. The study elucidates the significance of features and forecasting explainability through SHAP (SHapley Additive exPlanations) values and Genetic Programming (GP) models. Furthermore, the study emphasizes the potential of counterfactual analysis in decision support, particularly in scenarios where users can modify indoor conditions like temperature to influence energy consumption. The significance of 'indoor temperature' emerged as a key feature. Overall, the research underscores the significance of model explainability in building energy management however the narrow scope of the population, and potential biases in the analysis suggest that further research is needed to validate these findings across different contexts.
Another research into "Explainable AI for predicting daily household energy usages" by Mohanty et al [43] comprehends the factors influencing household energy consumption which is crucial for stakeholder trust in smart city initiatives. The study presents valuable insights by integrating weather data, household characteristics, and energy usage from London during 2012-2013, highlighting how factors like temperature and daylight impact energy consumption. In this specific analysis, the Random Forest Regressor model was exclusively trained using daily energy usage as the independent variable, while all other parameters served as dependent variables. The model's performance was evaluated, yielding an MSE of 86.366, an RMSE of 9.2933, and an R2 score of 0.074. To gain understanding into the model's decision-making process, SHAP (SHapley Additive exPlanations) was used as the sole method. Despite these limitations, the paper contributes to the field by addressing the need for explainable AI in energy optimization, though a broader and more critical examination of existing research would enhance its impact.
Maarif et al's [41] study, focused on industry energy consumption, incorporates dis-

tinctive features such as current power factor, current reactive power, and load type. It offers various insights into the application of advanced machine learning techniques, particularly LSTM, for energy forecasting. Key variables identified for influencing energy usage are reactive power and time of day, but could benefit from a more critical analysis of how these factors vary across different industrial environments. Additionally, while the paper emphasizes the improved accuracy of LSTM over other models, it does not deeply explore potential limitations such as the computational cost of training LSTM models, especially in larger datasets or more complex industrial scenarios. Although their research explores a unique concept specifically for industries, incorporating their model evaluation process could be beneficial for this study.

The study by Pan et al [47] on "Data-driven estimation of building energy consumption with multi-source heterogeneous data," utilizing Seattle's building energy data deeply explains the forecasting of building energy usage and highlights the influential features through the CatBoost model. A key strength of the paper is its focus on the model's ability to handle categorical variables and its performance in predicting energy use intensity compared to traditional methods like Random Forest(RF) and Gradient Boosting Decision Trees. However, it's worth noting that the study does not take into account external factors such as temperature, wind speed, and weather in its analysis and does not delve deeply into the potential limitations of this model, such as computational complexity or the need for extensive parameter tuning.

The research on "eXplainable AI (XAI)-Based Input Variable Selection Methodology for Forecasting Energy Consumption" by Sim et al [54] introduced an XAI-based methodology for selecting input variables in energy consumption prediction. Gas consumption data from a diverse 17-story building, including commercial properties and offices, were examined over one year. The study leveraged SHAP method to analyze the impact of each input variable on model predictions, facilitating the selection of optimal variables. These included information regarding time, past energy consumption and climate data. Based on these variables, energy consumption forecasting models were evaluated and noteworthy results were obtained. The analysis categorized input variables into Weak, Ambiguous and Strong groups, revealing the effectiveness of the combination of Strong and Ambiguous variables in enhancing model performance. The research concluded that utilizing high-impact variables identified through XAI analysis significantly improved energy forecasting models. Futhermore, the research acknowledges its limited scope due to the focus on a specific building and a limited set of input variables, encouraging future studies to investigate the inclusion of socioeconomic variables for various building types.

Woong C. et al's [16] research on "Analysis of input parameters for deep learning-based load prediction for office buildings in different climate zones using eXplainable Artificial Intelligence" underscores the need for a substantial amount of sensor data while emphasizing the significance of reducing the associated sensing and preprocessing costs to encourage the adoption of predictive building energy con-

trol systems. To achieve this, the research evaluates the relative importance of input variables using both global sensitivity analysis (SRC) and explainable AI (XAI) techniques like LIME and SHAP. Notably, the study reveals that, assisted by XAI techniques, accurate deep learning models can be constructed with fewer input variables compared to the global sensitivity analysis method. SHAP emerges as superior to LIME in retaining accuracy with fewer essential input variables, making it a preferred choice for building cost-effective deep learning-based building load prediction systems. The research also highlights the varying impact of input parameters, such as outdoor temperature, solar radiation, and time variables, depending on the specific climate zone, which underscores the importance of localized climate considerations in model development. However, the study does have some limitations, particularly in its reliance on simulated data rather than real-world empirical data, which may introduce uncertainties.

The research on "Toward explainable and interpretable building energy modelling: an explainable artificial intelligence approach" by Zhang et al [59] in the domain of building energy modeling, quantifies the impact of features and employs Partial Dependence Plots (PDP) for detailed explanations by the development of model-agnostic explanation and interpretation modules with a reference building energy model derived from a substantial dataset. By applying these techniques, the authors provide insights into how different features, such as floor area and building ID, impact energy consumption, revealing that the relationship between features and energy usage is not always linear or constant. For example, Building ID distinguishes buildings based on their structural composition, such as fireproofed steel or reinforced concrete frames, which influences energy efficiency differently across building types. This granular understanding is crucial for optimizing energy systems and making informed decisions in grid management. Noteworthy findings reveal that feature importance dynamically changes with varying feature values, providing nuanced insights. The use of surrogate decision trees for interpreting complex models further demonstrates how localized model interpretations can be both accurate and intuitive, making the results more accessible to system operators and building managers. It's crucial to note that the paper acknowledges the complexity of global building energy models, and while localized interpretations with local surrogates are realistic and accurate, there may be challenges in capturing all intricacies.

The research on "Scenario-based prediction of climate change impacts on building cooling energy consumption with explainable artificial intelligence" by Debaditya C. et al [12] addresses a critical gap in the field of building energy modeling, predicting continuous daily building energy consumption for space cooling amid climate change using an innovative eXplainable Artificial Intelligence (XAI) model. Integrating XGBoost and SHAP, the model aligns with CMIP6-SSP scenarios, overcoming limitations of existing morphing-based weather generators. Key findings highlight a sharp rise in energy consumption after the year 2050, underscoring the critical need for sustainable pathways moving forward. Regional variations in en-

ergy savings potential under climate change scenarios are identified, with hotter climates showing higher absolute savings potential Despite offering valuable insights, the study acknowledges challenges in long-term predictions and the model's interpretability, influenced by climate complexities and socioeconomic uncertainties.
The study by Zhang et al [60] on Seattle's building energy performance data, innovatively incorporated urban morphology and building geometry considerations into assessing building energy performance and greenhouse gas (GHG) emissions. Using LightGBM and XAI (SHAP), it proved that total gross floor area (GFA) and natural gas are crucial factors affecting site energy use intensity and GHG emissions. The LightGBM method outperformed XGBoost, Random Forest, and Support Vector Regression, offering a more accurate solution for urban planning. Unlike traditional models, this approach allows for a more accurate estimation of energy use by considering factors such as urban morphology, building geometry, and physical features. By improving prediction accuracy by 33.46% over models that only consider building characteristics, this study makes a significant contribution to energy efficiency planning in urban environments, highlighting the critical role of building and urban design in sustainable energy management. However, the study acknowledges limitations, suggesting future work on optimization strategies, multi-objective optimization, and extending the approach to different building types and cities.
Another study by Wenninger et al [56] found that the QLattice algorithm, while slightly less accurate than traditional models like Extreme Gradient Boosting (XGB) and Multiple Linear Regression (MLR), offers a unique advantage in terms of explainability. The QLattice's ability to generate simple mathematical expressions allows for easier interpretation of the factors influencing building energy consumption, bridging the gap between complex machine learning models and user-friendly insights. Additionally, the algorithm showed competitive performance in prediction accuracy, though it ranked slightly behind the best-performing models. The most influential factors were the size of the living space, which was the most significant determinant, and the type of energy source used, such as gas or oil. Additionally, the thickness of the outer wall insulation and the type of window glazing also played crucial roles in influencing energy efficiency. The study noted some limitations, including the relatively long training times required by the QLattice and the potential need for further validation across different datasets and building types to fully assess its generalizability. Despite these limitations, the QLattice presents a promising tool for integrating explainable AI into energy performance predictions, potentially increasing trust and adoption in real-world applications.
The study by Fan et al [20] presents a novel methodology for explaining and evaluating data-driven building energy performance models using interpretable machine learning techniques. The study demonstrates that the use of Local Interpretable Model-Agnostic Explanations (LIME) allows for detailed local explanations of individual predictions, offering insights into the inference mechanisms of the models. The study also introduces a novel "trust" metric to evaluate the reliability of individ-

ual predictions. This metric considers both the number and strength of supporting evidences for each prediction, providing a more nuanced assessment of model performance beyond traditional accuracy metrics. The findings suggest that models like Support Vector Machines (SVM) and Multi-Layer Perceptrons (MLP) produce more trustworthy predictions compared to tree-based models like Random Forests (RF) and Extreme Gradient Boosting (XGB), despite the latter showing higher overall accuracy. However, the study has some limitations, including the potential for low trust values even with acceptable model accuracy, which may complicate the interpretation of results. Additionally, while the methodology enhances interpretability, it still relies on complex statistical techniques that may require further simplification or validation across different datasets to ensure broader applicability.

2.2 Gaps in Existing Research

The current body of literature on Explainable Artificial Intelligence (XAI) in energy usage prediction highlights several critical research gaps that need to be addressed. First, there is a significant gap in the generalizability of XAI models across diverse contexts. Studies like those by Sakkas et al [53] have focused on specific building types or geographic regions, which limits the applicability of their findings to other settings. This lack of broader validation suggests a need for future research to develop XAI models that can be effectively applied across a variety of building types, climates, and geographic regions, ensuring that the insights derived are universally relevant.

Another gap is the limited integration of a broader range of environmental factors in energy consumption models. For instance, while Mohanty et al [43] have integrated weather data and household characteristics into their models, their focus has been relatively narrow, considering only specific variables like temperature and daylight. There is a need to include a wider array of environmental factors—such as wind speed, humidity, and seasonal variations—in these models to enhance their predictive accuracy and generalizability across different climates.

Additionally, the computational efficiency and scalability of advanced machine learning models, such as those used by Maarif et al [41], represent another critical research gap. Although these models have shown improved accuracy, their high computational demands limit their practical application, especially in large-scale industrial settings. Research is needed to develop computationally efficient models that can scale effectively without compromising accuracy, making XAI more accessible and practical for broader applications.

Furthermore, there is a gap in the development of multi-objective optimization frameworks in energy forecasting models. Woong C. et al [16] have highlighted the need for reducing sensing and preprocessing costs in predictive building energy control systems, but there has been little exploration of strategies that balance these costs with predictive accuracy. Future research should focus on optimizing for multiple objectives, such as minimizing data collection costs while maintaining high

predictive accuracy, to make these systems more economically viable.

The challenges of long-term predictive accuracy amid climate change, as explored by Debaditya C. et al [12], also reveal a research gap. The uncertainties inherent in long-term climate projections and the complexities of forecasting over extended periods require the development of more robust models. These models should account for a wider range of climate scenarios and socioeconomic factors to improve the reliability of long-term energy forecasts.

Finally, enhancing the interpretability and trust in AI models remains a significant gap. While Fan et al [20] introduced a trust metric to evaluate the reliability of AI predictions, the complexity of interpreting these metrics poses a challenge for practical application. There is a need to develop more intuitive and accessible interpretability frameworks that can build trust among non-expert users, thereby increasing the adoption of AI-driven energy management solutions.

2.3 Research Questions

Building upon the identified research gaps, this section outlines the key research questions that will guide the subsequent investigation. The purpose of these questions is to address the limitations in the existing body of literature and to explore new avenues that have not been thoroughly examined. The research questions are designed to ensure that the study contributes to both theoretical understanding and practical applications in the field of Explainable Artificial Intelligence (XAI) for energy consumption forecasting.

Specifically, these questions aim to investigate the generalizability of XAI models across diverse contexts, the integration of a broader range of environmental factors, the development of computationally efficient and scalable models, the application of multi-objective optimization in energy forecasting, the enhancement of long-term predictive accuracy amid climate change, and the improvement of interpretability frameworks to build trust among users. By addressing these questions, this research seeks to fill the gaps in the current literature and provide actionable insights that can be applied in real-world settings. The following research questions are proposed:

Firstly, most studies fail to integrate heating and electrical energy consumption, hindering accurate assessments of Total Energy Usage Intensity (EUI). This disregards the significant interdependencies between these energy sources, leading to limited forecasting.

Secondly, many existing approaches neglect the impact of external factors like temperature, wind speed, and weather conditions on energy consumption. Ignoring these crucial influences drastically restricts the generalizability and predictive power of forecasting models.

Thirdly, the limited scope of explainability and reliance on single Explainable AI technique for interpretation limits the comprehensiveness of understanding energy consumption patterns. This narrow focus can obscure valuable insights, hindering effective energy management strategies.

Following a comprehensive review of previous research and identifying the research gap, this study aims to achieve the following objectives.

- **RO1:** Assess the accuracy of ML models for overall energy prediction using both heating and electrical data, while considering external factors along with the building characteristics, comparing their performance and gaining interpretability through XAI techniques.
- **RO2:** What are the most influential factors in Building Energy Consumption according to XAI models? This focuses on using feature attribution methods in XAI to identify key variables affecting energy consumption in buildings. For example, different building characteristics like size, age, insulation quality or the external factors like weather, wind speed etc may be linked with energy usage, under the interpretation provided by XAI.
- **RO3:** How can XAI techniques improve the interpretability of Energy Usage prediction Models for non-expert users? This question investigates how XAI can make Energy Usage prediction models more understandable for users without technical expertise, thereby enhancing decision-making and policy formulation.

3 Theoretical Background

The advent of Machine Learning (ML) in the energy sector has revolutionized how energy consumption patterns are analyzed and predicted. ML offers a powerful toolkit for creating models that capture complex, non-linear relationships between multiple variables, thus allowing for more accurate predictions in comparison to traditional statistical methods. These predictions are critical in optimizing energy usage, managing grid loads, and enabling sustainability initiatives in response to growing global energy demands.

3.1 Overview of ML Algorithms in the Energy Industry

Traditionally, statistical models such as linear regression and autoregressive integrated moving average (ARIMA) were used for energy demand forecasting [58]. While useful in specific scenarios, these models struggle with the high complexity of modern energy systems, which are influenced by numerous interconnected factors, including weather conditions, socio-economic activity, and regional infrastructure. These traditional methods are limited in capturing non-linear interactions and often underperform when handling large, diverse datasets [28].

Machine learning approaches, particularly ensemble methods like Random Forest (RF), XGBoost, and CatBoost, have emerged as robust alternatives, offering significant improvements in accuracy and flexibility. A study by Ahmad et al [3] on comparison of Random Forests and Neural Networks for estimating building energy use showed that Neural Networks outperformed Random Forests by a small margin. Nonetheless, any missing data in the application may be handled by the Random Forest models with ease. Therefore, even with part of the input values missing, the Random Forests were still able to make correct predictions. Another research by Cao et al [10] proved random forest and gradient boosting (XGBoost) algorithms better than linear regression, lasso regression and ridge regression. Hence the decision based ML models were chosen in this research for the prediction task. Below, these methods are discussed in detail, focusing on their applicability to energy usage prediction.

3.1.1 Random Forest (RF)

Developed by Breiman [9] Random Forest is a well-established ensemble learning method that constructs multiple decision trees during training, and combines their results to improve predictive performance. Each tree $T_b(x)$ in the forest is built from

a random subset of the training data, and the final prediction is obtained by averaging (in regression tasks) or voting (in classification tasks). This ensemble approach reduces the risk of overfitting, which is a common issue with individual decision trees, and increases the overall robustness of the model [9].

Random Forest Formula:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^{B} T_b(x)$$

Here, B is the total number of trees, and $T_b(x)$ is the prediction of the b-th decision tree for input x. Random Forest averages the predictions of all trees, reducing overfitting and improving generalization by combining multiple weak learners [9].

In energy usage prediction, RF is particularly effective because it can handle high-dimensional data and does not require extensive data preprocessing. RF is also robust to outliers and missing data, both of which are common in real-world energy datasets. The majority of literature forecasted home energy (heat and electricity) usage having weather, calendar, and demographic data using decision tree methods [42]. Random forests are known for their robustness, accuracy, and ability to handle a variety of data types and their inherent feature importance mechanism offers a degree of interpretability [3].

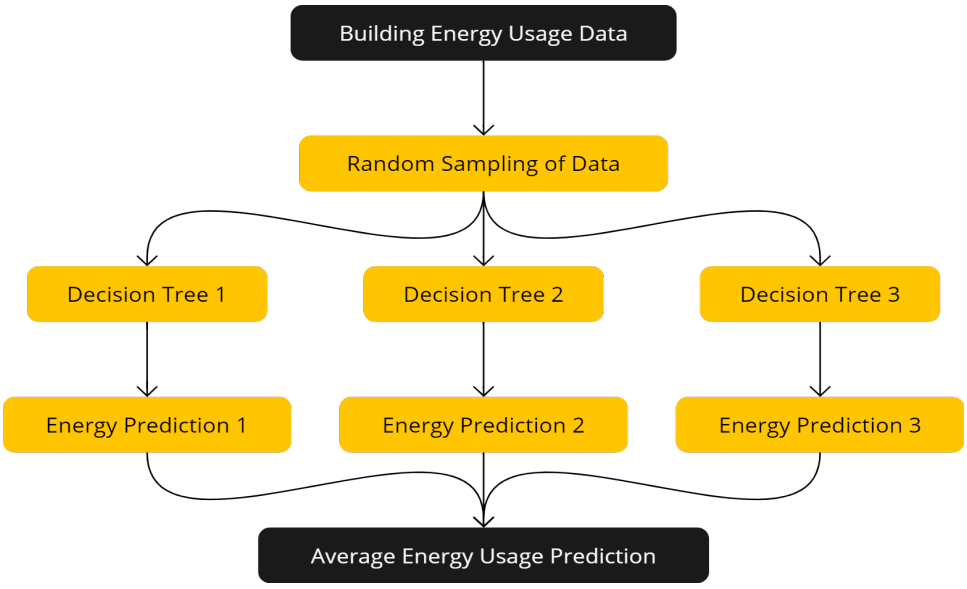


Figure 3.1: Random Forest

3.1.2 XGBoost

XGBoost, short for eXtreme Gradient Boosting, is a scalable, distributed gradient boosting algorithm designed to optimize both speed and accuracy [14]. Unlike traditional boosting methods, XGBoost introduces techniques like regularization and early stopping to prevent overfitting, making it ideal for handling large-scale, noisy energy datasets. Its ability to leverage parallel processing and its flexibility in handling missing values and categorical features contribute to its popularity in energy prediction tasks [18].

The prediction of XGBoost for a given input x is [14]:

$$\hat{y}_i = \sum_{t=1}^{T} F_t(x_i)$$

where T is the number of trees, and $F_t(x_i)$ is the prediction of the t-th tree for instance x_i. The model minimizes the following regularized objective:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(F_t)$$

where $\ell(y_i, \hat{y}_i)$ is a differentiable loss function (e.g., squared error for regression), and $\Omega(F_t) = \gamma T + \frac{1}{2}\lambda \|w\|^2$ is a regularization term that penalizes the complexity of the model, with γ controlling the number of leaves, and λ controlling the leaf weights w.
The evaluation study for assessing Brazil's energy consumption by Leme et al [36] confirmed that the Gradient Boosting predictive model performed the best, outperforming the other approaches in terms of accuracy.

Another research by Ravinder et al [40] showed that XGBoost outperformed LSTM and GRU algorithms in phrases of accuracy, in shape to the data, and predictive power. The smaller errors, better accuracy, and robust in shape of the XGBoost method make it a beneficial preference for the energy utilization prediction task.

The study by Chakraborty et al [13] on a comprehensive comparison of machine learning techniques for building energy prediction found that XGBoost consistently outperformed other models, demonstrating superior accuracy and computational efficiency. This highlights XGBoost's potential as a valuable tool for accurately predicting building energy consumption in real-world applications.

3.1.3 CatBoost

CatBoost (Categorical Boosting) is a gradient boosting algorithm that improves upon standard gradient boosting by efficiently handling categorical features and reducing prediction shift via ordered boosting [50]. The prediction of CatBoost for a given input x is :

$$\hat{f}(x) = \sum_{t=1}^{T} \eta \cdot F_t(x)$$

Here, T is the total number of trees, η is the learning rate, and $F_t(x)$ is the prediction of the t-th tree. CatBoost applies ordered boosting, which uses only past data points to compute gradients at each iteration, preventing data leakage and improving generalization. It also employs target statistics to encode categorical features efficiently without overfitting [50].

This characteristic makes it highly suitable for energy prediction tasks where categorical variables, such as building type, location, and time of day, are prevalent. CatBoost's ability to automatically handle categorical variables leads to higher accuracy and efficiency compared to algorithms that require manual encoding [8]. A research by Bassi et al [8] proved the XGBoost model performed best in comparison LightGBM, and CatBoost on a synthetic building energy prediction dataset based on the three key metrics (RMSLE, RN-RMSE (%), R^2).

For example, in prediction of building energy consumption with multi-source heterogeneous data, CatBoost has shown to outperform other algorithms by effectively handling mixed types of features, including both continuous and categorical variables [47]. Its robustness and simplicity of use make it an attractive choice for industrial applications in energy forecasting.

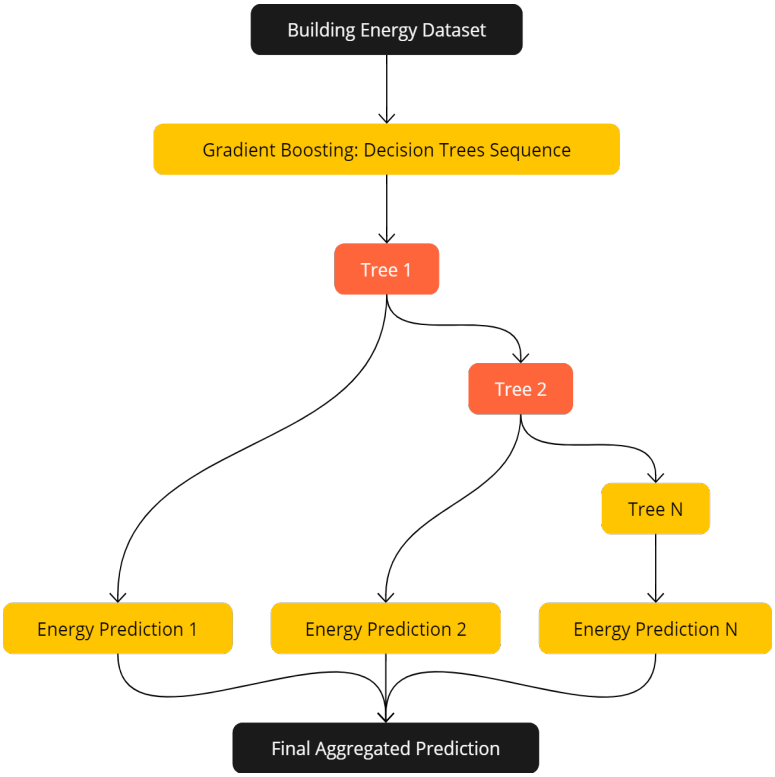


Figure 3.2: Catboost

3.2 Explainable AI (XAI) Algorithms for Energy Usage Prediction

While machine learning models like Random Forest, XGBoost, and CatBoost deliver high accuracy, their complexity often renders them opaque to end-users. In energy management, where decision-making involves critical infrastructure and large-scale investments, understanding why a model makes specific predictions is as important as the accuracy of the predictions themselves [7]. This need for transparency has led to the integration of Explainable AI (XAI) algorithms into energy usage prediction models.

XAI refers to a suite of techniques designed to make machine learning models more interpretable, thereby providing insights into how models arrive at their decisions. These insights are crucial for fostering trust, especially in industries like energy, where stakeholders demand explainability to validate the model's reliability and fairness [24]. XAI techniques empower stakeholders to understand the key factors influencing energy consumption, identify potential biases or limitations in the model, and make informed decisions regarding energy efficiency interventions. In the context of Site Energy Usage Intensity prediction, several XAI algorithms have

proven particularly valuable:

3.2.1 SHAP (SHapley Additive exPlanations)

SHAP is one of the most widely adopted XAI algorithms in the energy sector. Based on cooperative game theory, SHAP assigns a Shapley value to each feature, representing its contribution to the prediction [38]. SHAP is particularly useful in energy prediction because it provides both global explanations (overall feature importance across all predictions) and local explanations (feature importance for individual predictions).

SHAP values are calculated by considering all possible subsets S of features N, excluding feature i. For each subset, the marginal contribution of i is computed as $f(S \cup \{i\}) - f(S)$. The SHAP value is a weighted average of these contributions across all subsets, where the weight $\frac{|S|!(|N|-|S|-1)!}{|N|!}$ ensures fair distribution. The sum of SHAP values equals the difference between the model's prediction for an instance and the average model prediction [38].

SHAP Formula
The SHAP value for a feature i is computed using the following formula:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left[f(S \cup \{i\}) - f(S) \right]$$

Explanation:

- ϕ_i: SHAP value for feature i, showing its contribution.
- S**: Subset of features excluding i.
- $f(S \cup \{i\}) - f(S)$**: Marginal contribution of feature i to subset S.
- Weights: $\frac{|S|!(|N|-|S|-1)!}{|N|!}$ ensures fair distribution of contributions across all subsets.

In the research paper by Golafshani et al [22], titled "An Artificial Intelligence Framework for Predicting Operational Energy Consumption in Office Buildings", SHAP is utilized as a method for interpreting complex machine learning models. SHAP provides a comprehensive and consistent framework for explaining individual predictions by assigning each feature an importance value, based on its contribution to the model's output. The paper emphasizes the role of SHAP in identifying critical variables affecting energy consumption, such as occupancy, external temperature, and equipment usage, which are crucial for making informed decisions in building management. By using SHAP, the researchers were able to offer insights

into how specific features influenced the energy consumption predictions of their model, ensuring greater transparency and trust in the machine learning outputs.

One of the key advantages of SHAP, as highlighted in this study, is its ability to offer both global and local explanations. This allows building operators and energy managers not only to understand the general importance of each variable across all predictions but also to drill down into individual instances to see why a particular energy consumption forecast was made. SHAP's foundation in cooperative game theory ensures that the feature attributions are fair and consistent across predictions, which is especially important in energy management systems where small changes in variables can have significant operational impacts. This study effectively bridges the gap between complex machine learning models and human understanding in fields like building energy forecasting.

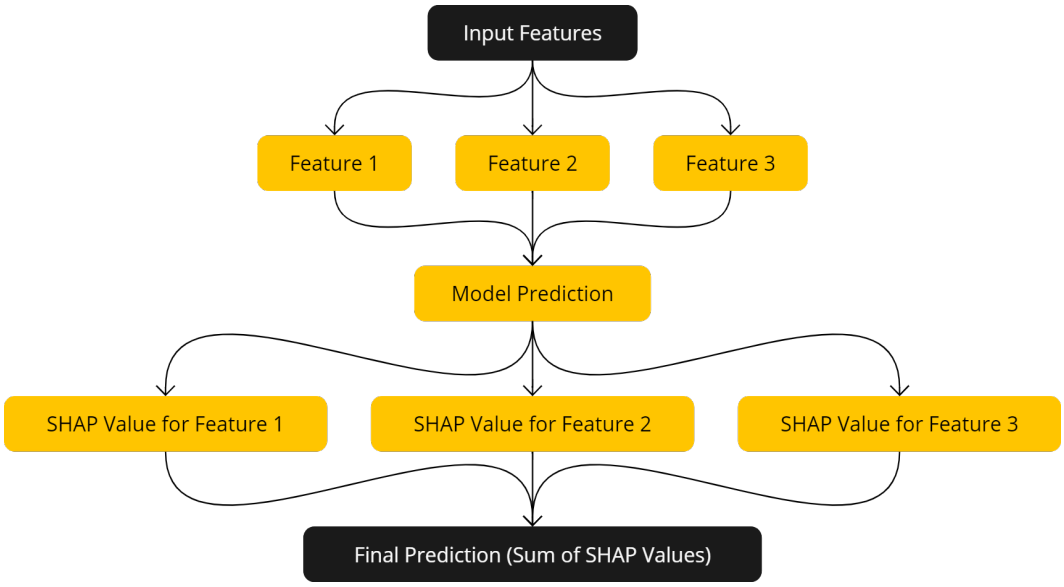


Figure 3.3: SHAP

3.2.2 LIME (Local Interpretable Model-agnostic Explanations)

LIME offers another approach to model interpretability by approximating a complex model with a simpler, interpretable model, such as a linear model or decision tree, for individual predictions [52]. LIME's flexibility allows it to be applied across a wide range of machine learning models, making it a versatile choice for interpreting energy usage predictions.

LIME explains model predictions by fitting an interpretable model $g(z)$ locally around the instance x being explained. It optimizes the following:

$$\xi = \arg\min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

Here, f is the original model, g is the interpretable local model (e.g., linear), $\mathcal{L}(f, g, \pi_x)$ measures how close g's predictions are to f's in the locality defined by the kernel function π_x around x, and $\Omega(g)$ penalizes complexity of g. LIME perturbs the input data to generate samples and then fits g to these samples weighted by π_x, focusing on local fidelity [52].

In A Survey of Methods for Explaining Black Box Models by Guidotti et al [23], LIME is introduced as one of the key tools for locally interpreting complex machine learning models. LIME operates by building a simplified, interpretable model (such as a linear model) around a particular instance or prediction of interest. It achieves this by perturbing the input data around that instance, generating synthetic examples, and observing how the black-box model responds. This localized explanation offers insights into why the model made a specific prediction, without needing to understand the entire global behavior of the model. The method is highly flexible since it can be applied to any kind of model, regardless of its architecture, making it popular for explaining black-box models like neural networks and gradient-boosting models.

The paper also highlights some trade-offs when using LIME. One notable drawback is its reliance on synthetic data perturbation, which can sometimes lead to instability in explanations, particularly in regions where the model's decision boundary is complex. Moreover, LIME's approach can be computationally intensive because it requires generating and processing several new instances around the input to probe the model's behavior. Despite these limitations, LIME remains widely used due to its model-agnostic nature and effectiveness in generating local explanations that are easily interpretable by non-expert users. This has made it particularly useful in domains such as healthcare, finance, and legal applications, where understanding individual predictions is critical for trust and decision-making.

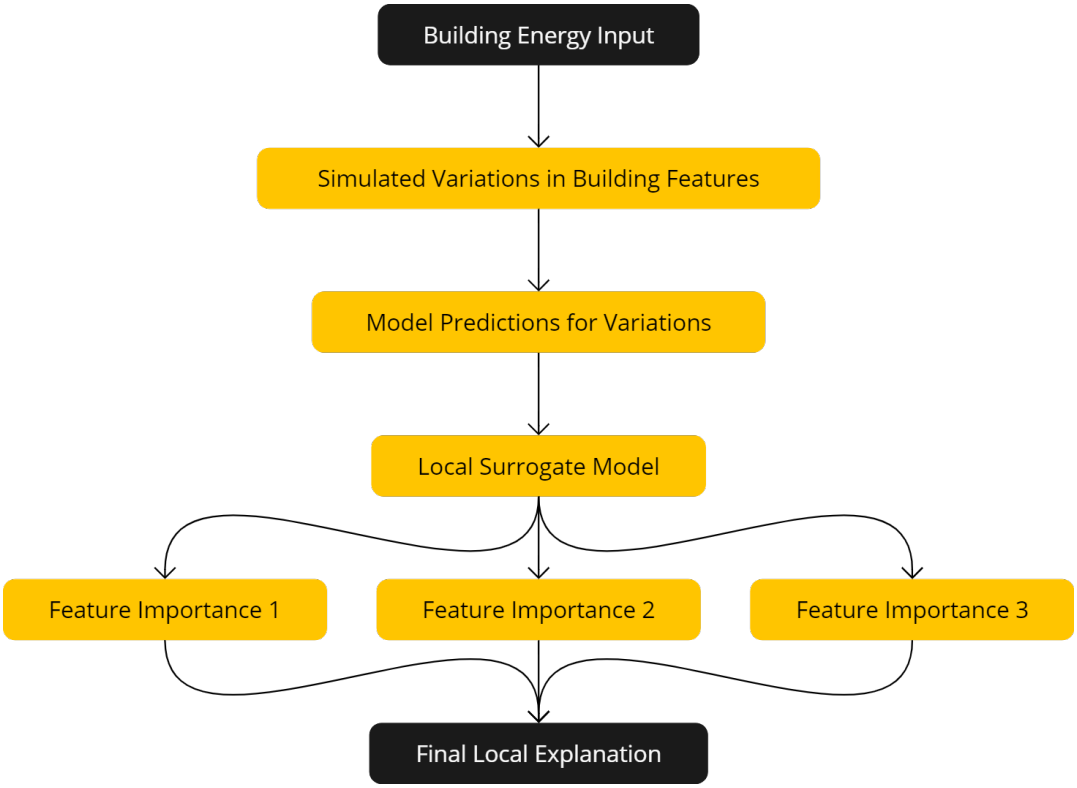


Figure 3.4: LIME

3.3 Evaluation

3.3.1 ML Evaluation Metrics

Model performance was rigorously evaluated using established metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of determination (R^2), providing a quantitative assessment of the predictive accuracy of each model on the unseen test set.

RMSE, a widely adopted measure of the discrepancies between predicted and actual values, quantifies the magnitude of errors and is particularly sensitive to large deviations [30].

The Root Mean Square Error (RMSE) is calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where y_i represents the actual values, $\hat{y}_i$ represents the predicted values, and n is the number of observations [26].

MAE, on the other hand, provides a straightforward average of the absolute errors, offering a more interpretable measure of prediction accuracy [57].

The Mean Absolute Error (MAE) is calculated as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where y_i represents the actual values, $\hat{y}_i$ represents the predicted values, and n is the number of observations [51].

Lastly, R^2, also known as the coefficient of determination, indicates the proportion of variance in the target variable explained by the model, serving as a comprehensive measure of overall model fit [34].

The Coefficient of Determination (R^2) is calculated using the formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where y_i are the actual values, $\hat{y}_i$ are the predicted values, $\bar{y}$ is the mean of the actual values, and n is the number of observations [15].

3.3.2 XAI Evaluation Metrics

According to Zhou et al [61], to determine if an application's explainability is achieved, one could compare the available explanation methods and identify the preferred explanations from the comparison. The research "Evaluating the quality of machine learning explanations: A survey on methods and metrics" [61] explored the numerous explanation methods presently available and categorized in two main categories as; "Human - Centred Evaluation (HCE)" and "Functionality - Grounded Explanation (FGE)". While HCE offer valuable insights by directly involving users, they rely on subjective measures like trust and confidence, making it difficult to compare evaluation results objectively. Despite numerous HCE studies, there's a lack of standardized approaches, particularly in experimental design and selection of subjective measures. This inconsistency hinders comparing the quality and effectiveness of different evaluations. On the other hand FGE techniques focus on evaluating the effectiveness of explanations for machine learning models based on their functionality and their alignment with the task or application at hand. Unlike HCE which relies on user perception and trust, FGE techniques leverage quantitative metrics and domain knowledge for assessment. Therefore this study will utilise the Functionality - Grounded Explanation techniques.

1. **Task Fidelity:**

   In machine learning, task fidelity refers to how accurately a simpler, interpretable model approximates the predictions of a more complex, black-box model. Task fidelity is quantified as the proportion of times the local explanation model produces predictions that match or closely resemble the black-box

model's predictions [48].

Calculating Task Fidelity:
To compute task fidelity, the prediction from the local model (LIME) with the black-box model's actual predictions is compared [55].

Task Fidelity Formula:
The formula to calculate task fidelity is as follows:

$$\text{Task Fidelity} = \frac{\text{Number of Correct/Similar Predictions}}{\text{Total Number of Predictions}} \times 100$$

For a single instance, if the local explanation model's prediction (e.g. LIME) is close to the actual black-box model's prediction, the fidelity is said to be high for that instance.

2. **Feature Importance Consistency**
   - **Description:** This technique assesses the consistency between global and local feature importance explanations. It ensures that the key features identified as important at the global level (across the entire dataset) are also important at the local level (for individual predictions) [29].
   - **Evaluation Process:**
     - The global importance rankings, often derived from methods such as SHAP summary plots or feature importance scores, are compared with local instance-specific explanations generated by methods like LIME or individual SHAP plots.
     - Consistency between global and local explanations indicates that the model behaves predictably and that important features are recognized both globally and locally.
3. **Local vs Global Explanations Alignment**
   - **Description:** This technique examines the alignment between local explanations (which explain individual predictions) and global explanations (which explain overall model behavior).
   - **Evaluation Process:**
     - Local explanations from methods like LIME or SHAP for specific instances are compared with global explanation methods like SHAP summary plots or global feature importance rankings.
     - A strong alignment between local and global explanations suggests that the model's behavior is consistent and that the features influencing individual predictions are also significant on a broader scale, enhancing trust in the model [37].

4. **Generalizability**
    - **Description:** This technique evaluates how well the model's explanations generalize to unseen data or new instances. An effective XAI method should yield similar feature importance and explanations across different datasets (e.g., training vs. test sets) [17].
    - **Evaluation Process:**
        - The stability of feature importance or explanation patterns is analyzed across training and test datasets.
        - SHAP values, feature importance rankings, or consistency of local explanations in both datasets are examined to determine if the model's explanations generalize well.
        - If explanations remain consistent across different datasets, it indicates that the model's behavior is robust and suitable for real-world applications.

4 Methodology

This research aims to address the research questions by:

- Developing a comprehensive forecasting model that integrates heating and electrical energy consumption (EUI).
- Incorporating the influence of external factors, such as temperature, wind, into the forecasting process.
- Leveraging a combination of Explainable AI techniques to provide a richer and more nuanced understanding of energy consumption patterns.

This comprehensive approach is envisioned to lead to more accurate and informative EUI forecasts, ultimately enabling more effective energy management and optimization across diverse building types and climates.

4.1 Data

The data is collected over 7 years, in several states within the United States. It is created by Climate Change AI and Lawrence Berkeley National Laboratory and provided by WiDS Datathon [33].

4.1.1 Data Overview

There are 64 columns and about 80000 rows in the dataset, each corresponding to a different aspect of the building, weather, or energy-related information. It contain information about various commercial buildings, including their characteristics, energy efficiency, and environmental factors. It also includes weather-related data for the building locations, such as temperature, precipitation, and wind speed.

The various columns from the dataset are explained below as provided by Kaggle [33], the column names along with the data types are mentioned in appendix 6.1:

- Year_Factor: The year factor associated with the data.
- State_Factor: The state factor indicating the state associated with the data.
- Building_Class: The class of the building (e.g., Commercial).
- Facility_Type: The type of facility within the building (e.g., Grocery store or food market, Warehouse, Retail enclosed mall, Education other classroom).

- Floor_Area: The floor area of the building.
- Year_Built: The year the building was constructed.
- Energy_Star_Rating: The energy star rating of the building.
- Elevation: The elevation of the building location.
- Temperature Data (January to December): Minimum, average, and maximum temperatures for each month.
- Cooling_Degree_Days: Cooling degree day for a given day is the number of degrees where the daily average temperature exceeds 65 degrees Fahrenheit. Each month is summed to produce an annual total at the location of the building.
- Heating_Degree_Days: Heating degree day for a given day is the number of degrees where the daily average temperature falls under 65 degrees Fahrenheit. Each month is summed to produce an annual total at the location of the building.
- Precipitation_Inches: The amount of precipitation in inches.
- Snowfall_Inches: The amount of snowfall in inches.
- Snowdepth_Inches: The snow depth in inches.
- Avg_Temp: The average temperature over a year.
- Days_Below_30F, Days_Below_20F, Days_Below_10F, Days_Below_0F: Number of days below certain temperature thresholds.
- Days_Above_80F, Days_Above_90F, Days_Above_100F, Days_Above_110F: Number of days above certain temperature thresholds.
- Direction_Max_Wind_Speed: The direction of the maximum wind speed Given in 360-degree compass point directions (e.g. 360 = north, 180 = south, etc.).
- Direction_Peak_Wind_Speed: The direction of the peak wind speed.
- Max_Wind_Speed: The maximum wind speed.
- Days_With_Fog: Number of days with fog.
- Site_EUI: Site Energy Usage Intensity is the amount of heat and electricity consumed by a building as reflected in utility bills.
- ID: Building Id.

year_factor	state_factor	building_class	facility_type	floor_area	year_built	energy_star_rating	elevation	january_min_temp	january_avg_temp
7	State_1	Commercial	Grocery_store_or_food_market	28484.0	1994.0	37.0	2.4	38	50.596774
7	State_1	Commercial	Grocery_store_or_food_market	21906.0	1961.0	55.0	45.7	38	50.596774
7	State_1	Commercial	Grocery_store_or_food_market	16138.0	1950.0	1.0	59.1	38	50.596774
7	State_1	Commercial	Grocery_store_or_food_market	97422.0	1971.0	34.0	35.4	38	50.596774
7	State_1	Commercial	Grocery_store_or_food_market	61242.0	1942.0	35.0	1.8	38	50.596774

Figure 4.1: Data1

days_above_90F	days_above_100F	days_above_110F	direction_max_wind_speed	direction_peak_wind_speed	max_wind_speed	days_with_fog	site_eui	building_id
5	2	0	NaN	NaN	NaN	NaN	166.588554	75757
5	2	0	NaN	NaN	NaN	NaN	259.381565	75758
5	2	0	NaN	NaN	NaN	NaN	158.537090	75759
5	2	0	NaN	NaN	NaN	NaN	261.441520	75760
5	2	0	340.0	330.0	22.8	126.0	242.967711	75761

Figure 4.2: Data2

The feature space encompasses a diverse set of information. Features include building characteristics (e.g., floor area, year built), energy-related metrics (e.g., energy star rating, site EUI), weather data (temperature, precipitation, snowfall), and various other variables related to wind, fog, and days above or below specific temperature thresholds.

The dataset poses several challenges that warrant careful consideration in the analysis. Firstly, the dataset exhibits a high-dimensional feature space with 64 columns, introducing complexities in visualizing and interpreting relationships between variables. Secondly, the quality and completeness of the data need thorough scrutiny, as missing values or inaccuracies could compromise the reliability of analysis results. Furthermore, the inclusion of diverse variables, ranging from building characteristics and energy metrics to weather data, implies complex interactions that may necessitate advanced analytical methods. Lastly, the risk of overfitting is heightened due to the small sample size, where models may perform well on the available data but struggle to generalize to new instances. Addressing these challenges required meticulous data preprocessing, potential dimensionality reduction, and the application of appropriate analytical techniques, all while considering domain-specific knowledge for accurate interpretation of results.

4.1.2 Data Exploration

This section contains the highlights from the exploratory data analysis from the train and test datasets.

In the training dataset, there are six columns with missing values, namely: year_built, energy_star_rating, direction_max_wind_speed, direction_peak_wind_speed, max_wind_speed, and days_with_fog. On the other hand,

Table 4.1: Comparison of Training and Test Data Features

Feature	**Training Data**	**Test Data**
Number of observations	75757	9705
Number of columns	64	64
Integer columns	37	37
Float columns	24	24
Object columns	3	3
Duplicate observations	0	0

the test dataset contains two constant columns: year_factor and days_above_110F, and also has six columns with missing values, which are the same as those in the training dataset.

The initial focus of the exploratory data analysis was on identifying columns with near-constant values across the dataset. For instance, the feature *days_above_110F* showed minimal variability, holding the same value for approximately 99% of the training data and 100% of the test data. Similarly, the feature *days_above_100F* exhibited constant values in 95% of the training data and 90% of the test data. These columns were flagged due to their lack of variability, which could reduce their usefulness in model training and potentially introduce noise into the prediction process.

Further examination was conducted on columns with unrealistic or highly skewed values. The feature *direction_max_wind_speed* displayed a single value in 80% of the training observations. Similarly, the features *direction_peak_wind_speed* and *max_wind_speed* presented the same value for almost 80% of the training observations. These uniform distributions, which were also consistent across the test dataset, suggested that these columns contained isolated outliers or unrealistic values that may distort the results.

Additionally, anomalies were detected in the *year_built* feature. A total of 6 observations in the training set and 1 observation in the test set had a *year_built* value of 0 which is highly improbable and stands out as an outlier in both datasets. These anomalous values were replaced with NaN to maintain data integrity and ensure cleaner inputs for model training.

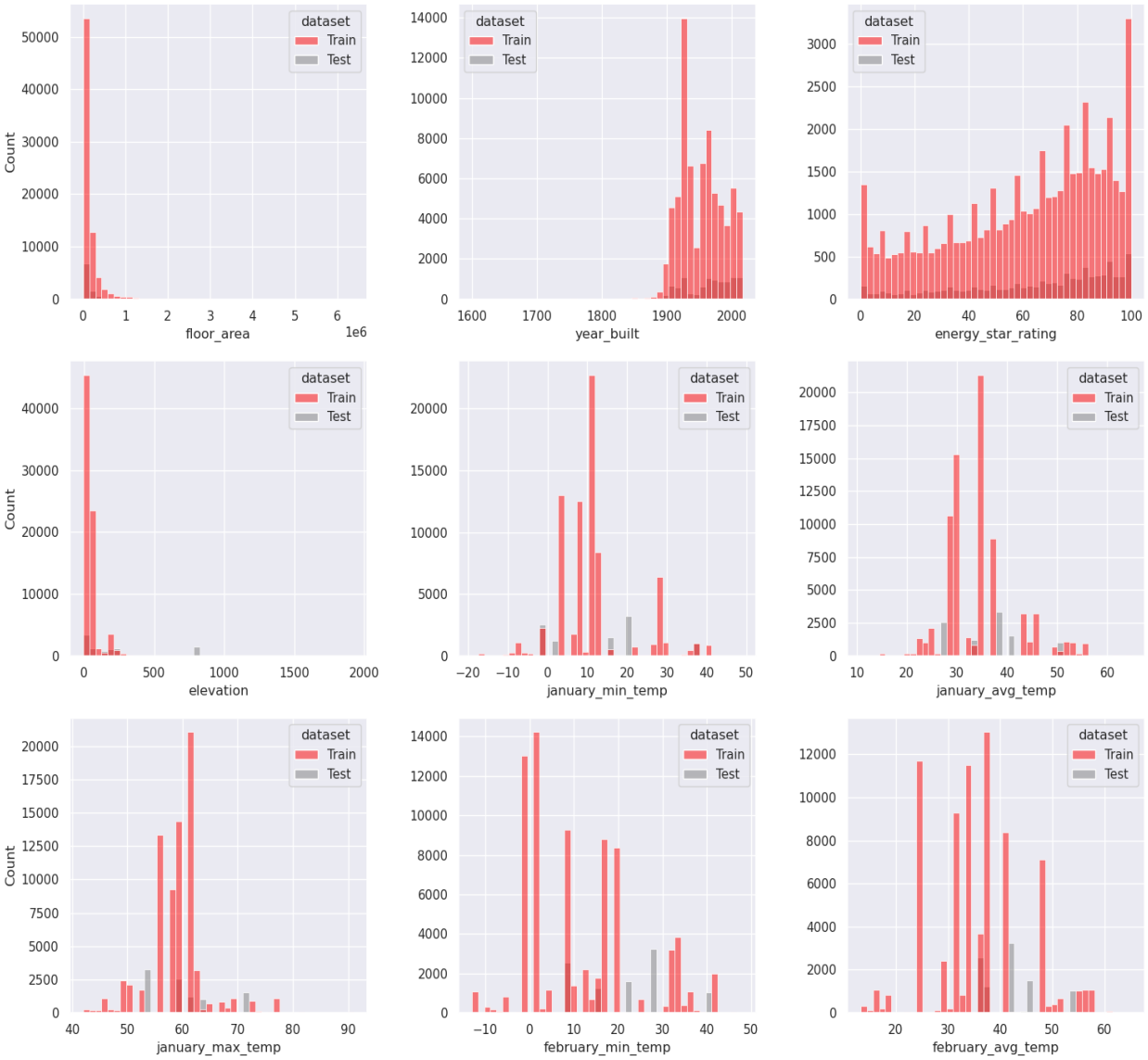


Figure 4.3: Histogram Part 1

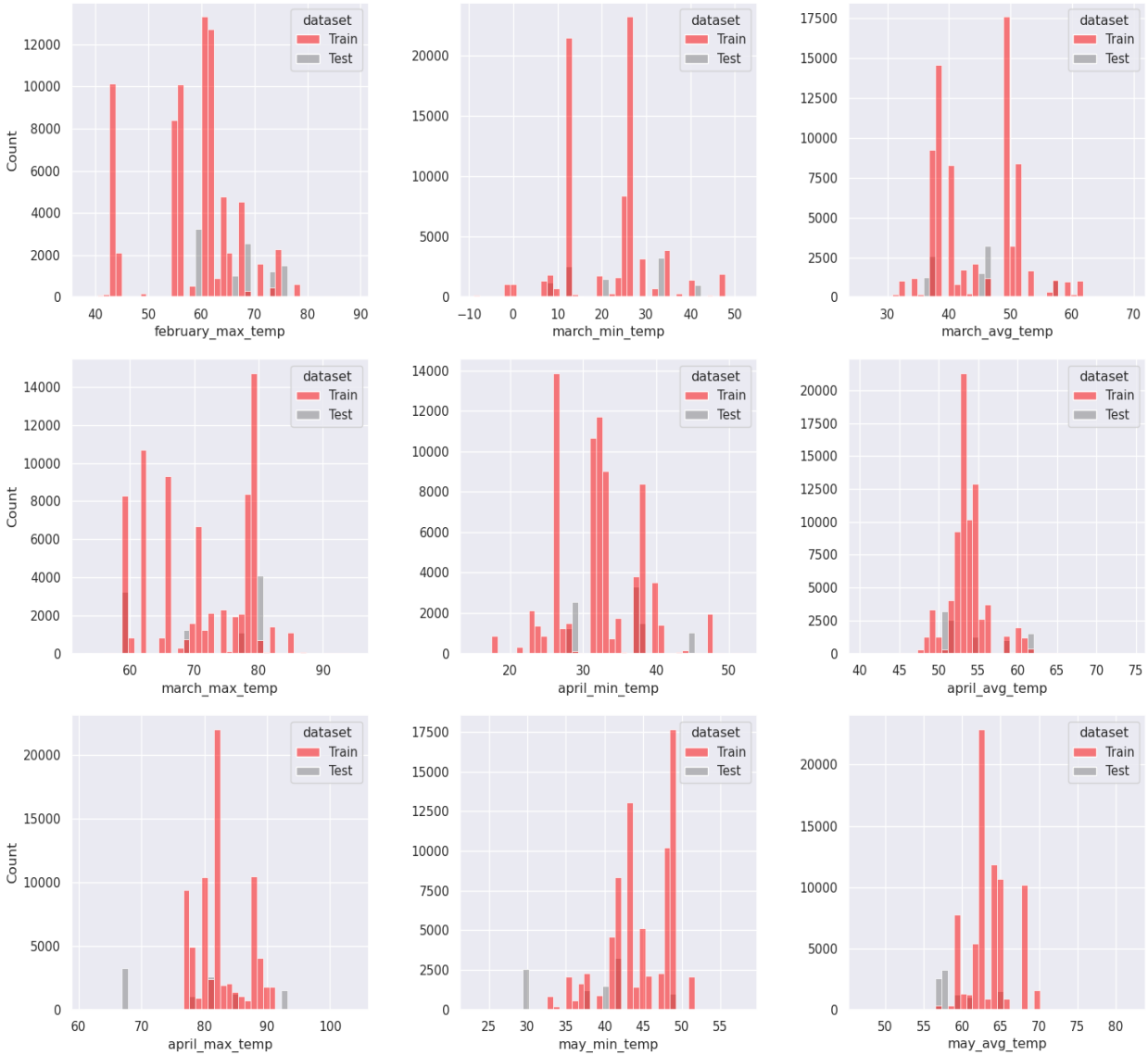


Figure 4.4: Histogram Part 2

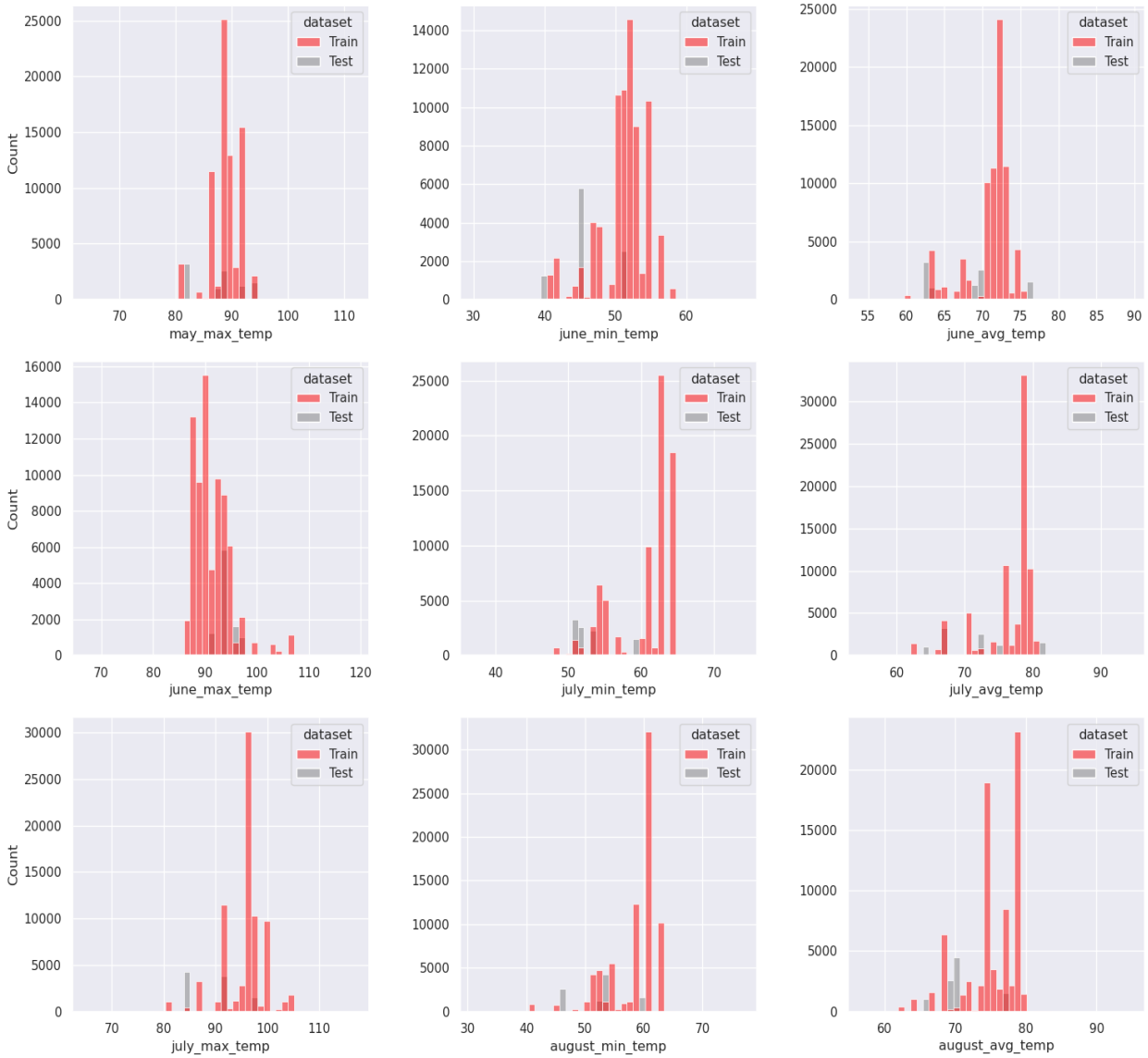


Figure 4.5: Histogram Part 3

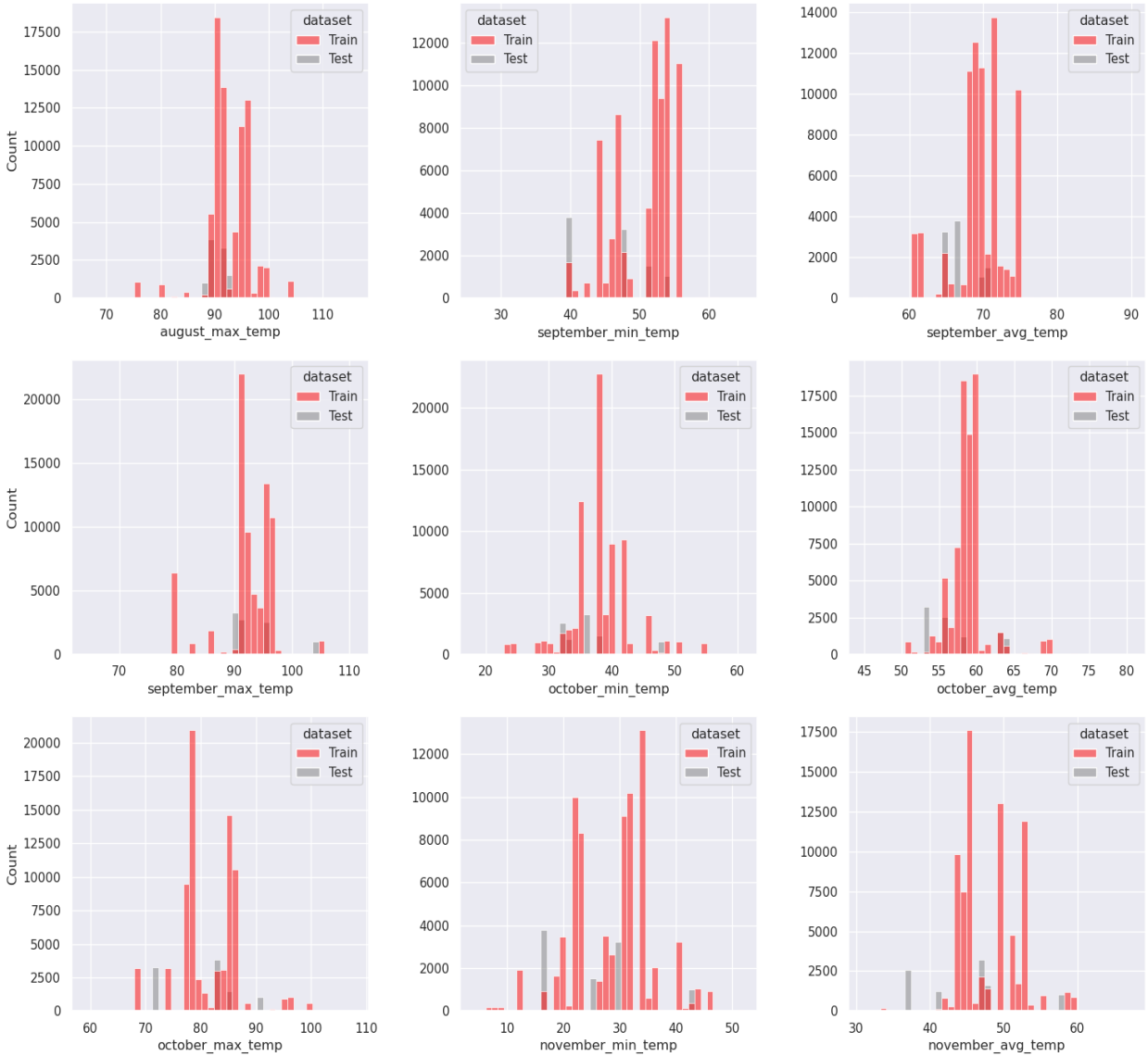


Figure 4.6: Histogram Part 4

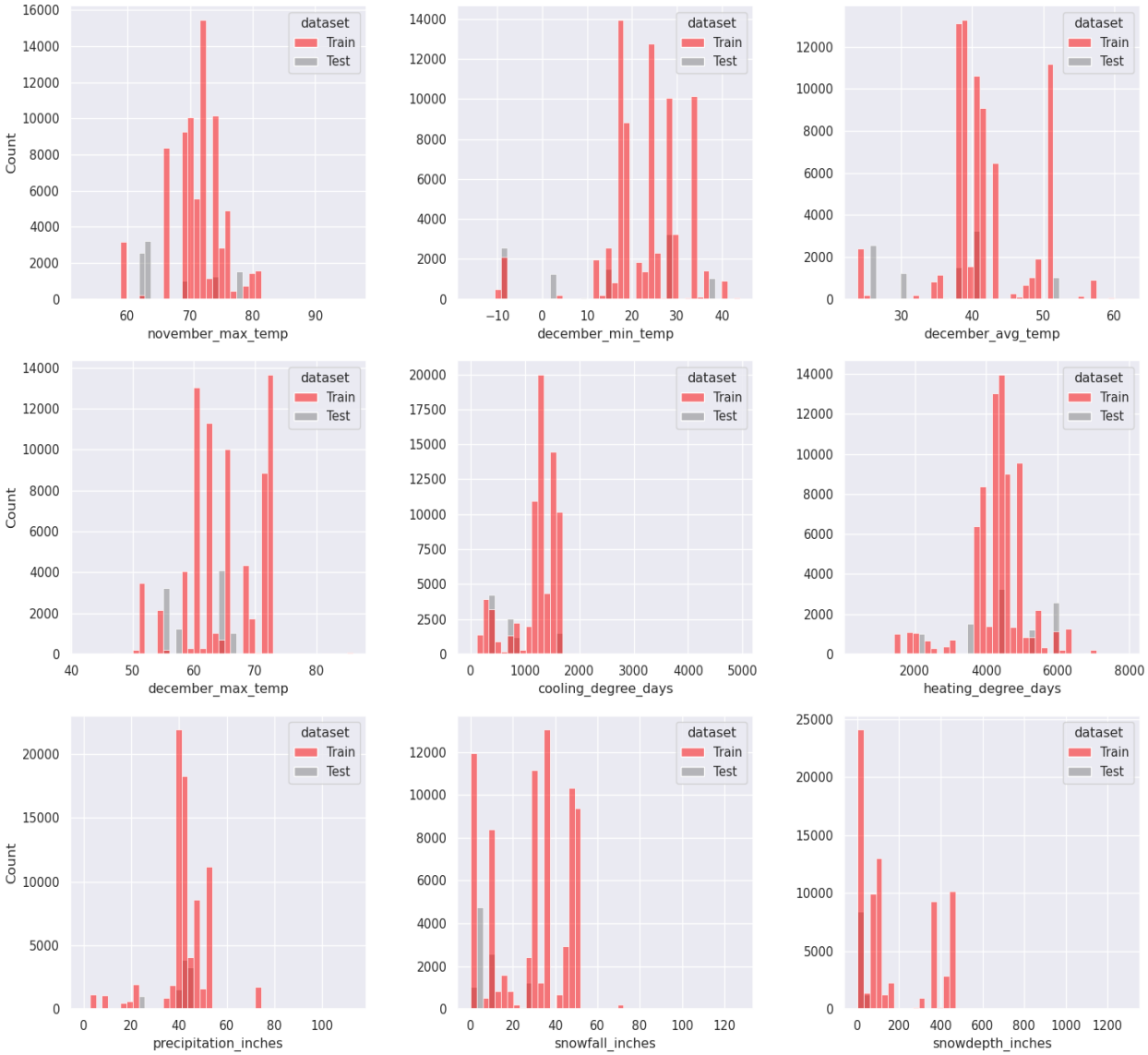


Figure 4.7: Histogram Part 5

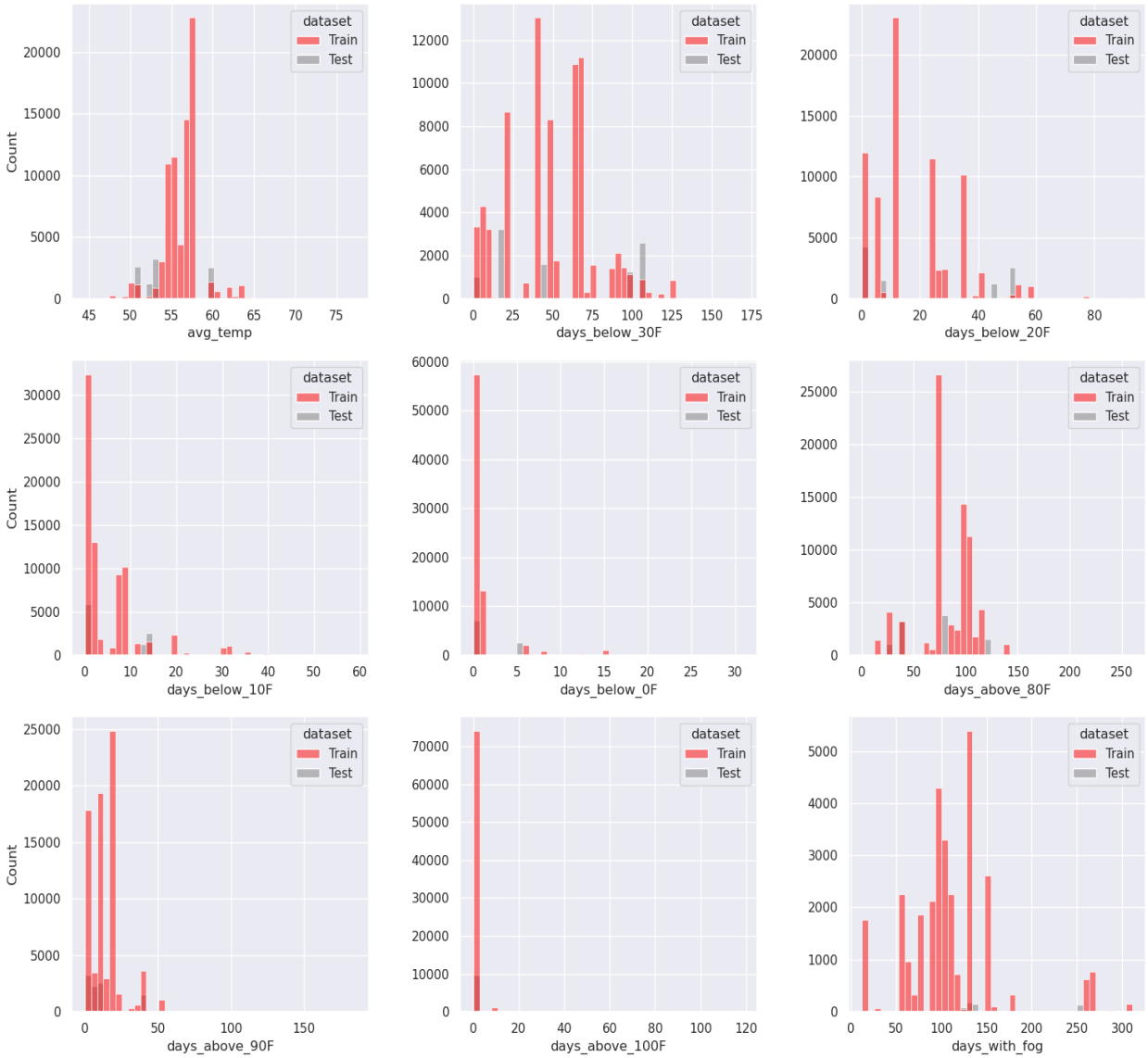


Figure 4.8: Histogram Part 6

From the above histograms in figures 4.3, 4.1.2, 4.5, 4.6, 4.7 and 4.8, it is found that the variables *floor_area* and *elevation* exhibit significant positive skewness, indicating that the majority of values are concentrated toward the lower end of the distribution, with a few extreme outliers. On the other hand, the variable *energy_star_rating* demonstrates moderate negative skewness, showing a higher concentration of values on the upper end.

The oldest building in the dataset was constructed in the year 1600, while the most recent building was completed in 2016. A large portion of the buildings in both

the training and test datasets were constructed after the year 1900. Additionally, a substantial number of buildings were built during the latter half of the 1920s and the early 1960s. It is also noticeable that the data shows sharp declines in the number of buildings constructed during the periods of both World War I (1914–1918) and World War II (1939–1945).

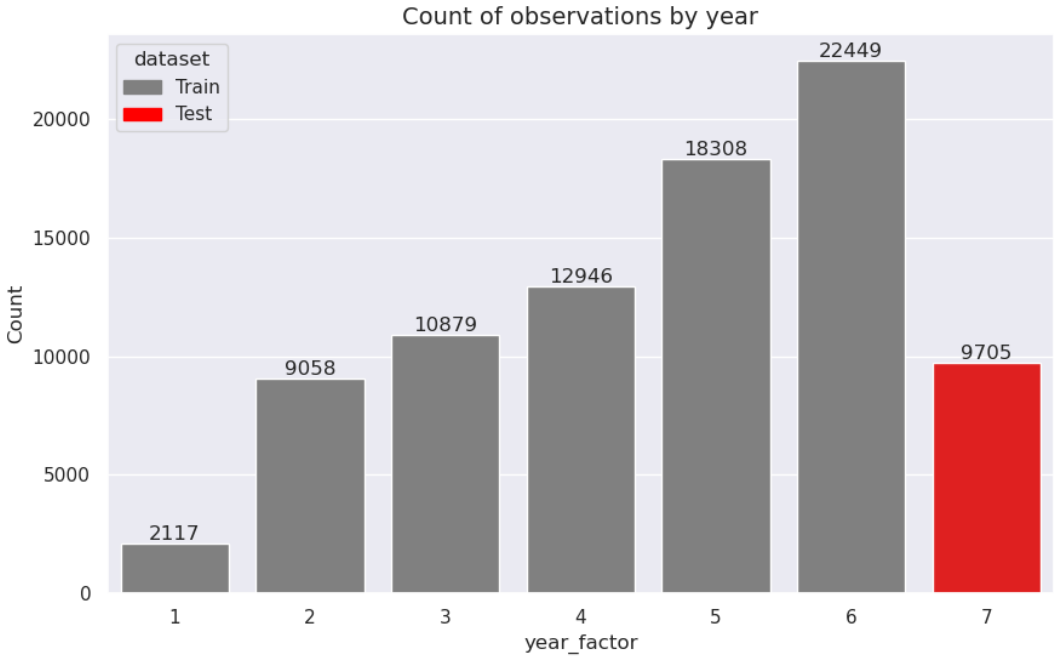


Figure 4.9: Frequency By Year

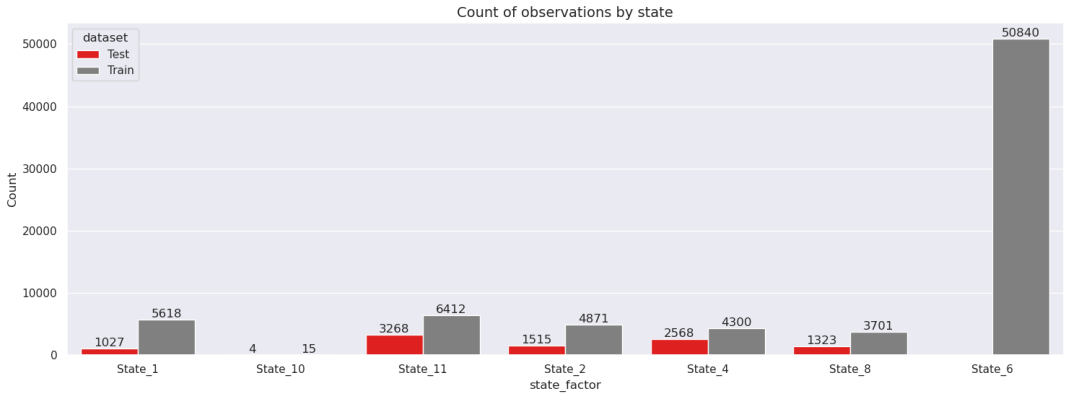


Figure 4.10: Frequency By State

The figures 4.9 and 4.10 show the count of observations by year and state respec-

tively in the training dataset and has the highest count in the year_factor 5 and state_factor 6 .

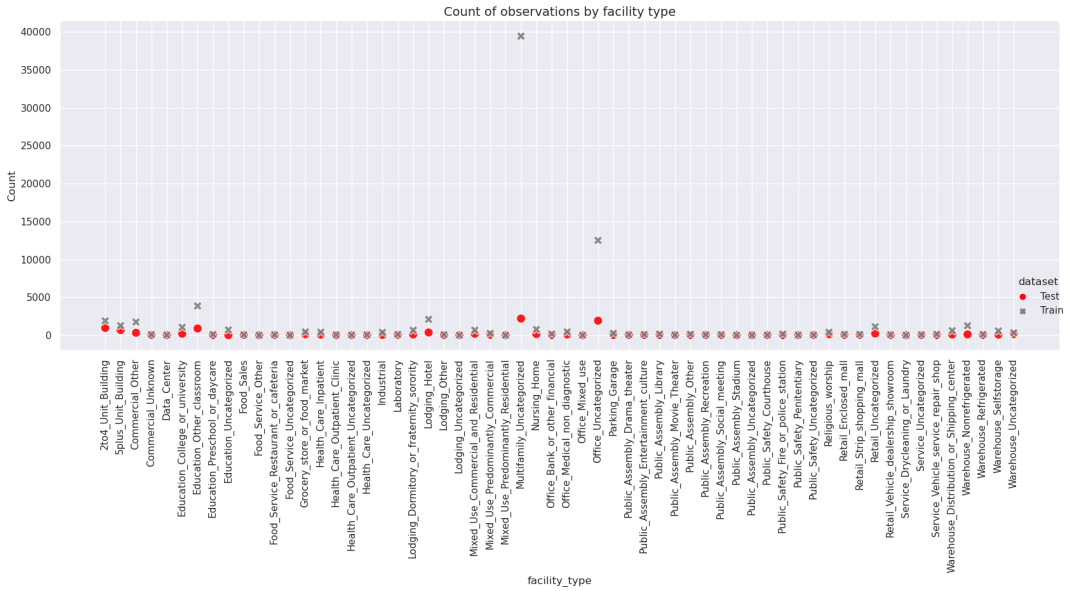


Figure 4.11: Frequency By Facility Type

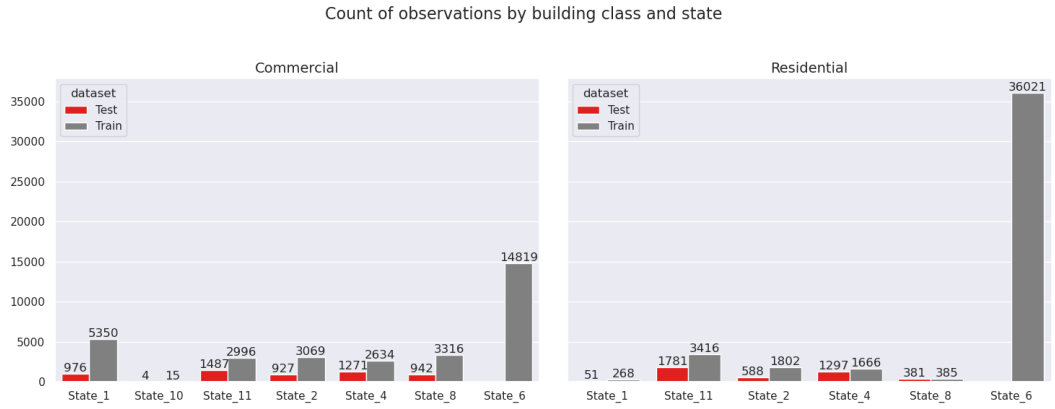


Figure 4.12: Frequency By Building Class and State

In the training dataset, more than 52% of the buildings are classified as *Multifamily_Uncategorized*, followed by 16% of the buildings categorized as *Office_Uncategorized*. In contrast, the test dataset displays a more even distribution, with *Multifamily_Uncategorized* and *Office_Uncategorized* representing 22.7% and 19.8% of the observations, respectively. Additionally the figure 4.12 shows that the data from *State_10* consists solely of commercial buildings.

These results indicate that the majority of buildings in both the training and test datasets are classified as either *Multifamily_Uncategorized* or *Office_Uncategorized*. It is clear that *Multifamily_Uncategorized* buildings are predominantly residential, whereas *Office_Uncategorized* buildings are primarily commercial in nature.

4.2 Implementation

This thesis investigated the prediction of Site Energy Usage Intensity (EUI) using machine learning algorithms and the interpretation of the resulting models using XAI techniques. The research process involved several key stages.

Initially, the data was split into training and test sets, followed by data exploration and feature understanding. After that cleaning and handling of missing values was performed on the training dataset. Subsequently, label encoding and feature scaling were applied to prepare the data for modeling. Regression models, specifically Random Forest, CatBoost, and XGBoost, were then employed to train and evaluate the models on the training set. The choice of regression analysis was motivated by its focus on predicting a single dependent variable based on a set of varying variables, making it ideal for forecasting tasks like EUI prediction.

While achieving accurate predictions was a primary objective, this thesis placed a particular emphasis on explainability. The subsequent phase of the research was dedicated to interpreting, explaining, and visualizing the models using various XAI (Explainable Artificial Intelligence) methods. These methods, including techniques like SHAP, LIME, and others, were applied to the best performing algorithm. The insights gained from these explanations were crucial in improving the interpretability of Energy Usage prediction Models for non-expert users.

The selection of specific machine learning algorithms was informed by a thorough exploratory data analysis, ensuring the chosen models were well-suited to the characteristics of the data and the research objectives. By combining predictive modeling with XAI techniques, this thesis aimed to create not only accurate but also transparent and interpretable models for SEUI prediction, fostering trust and enabling informed decision-making in the realm of energy management.

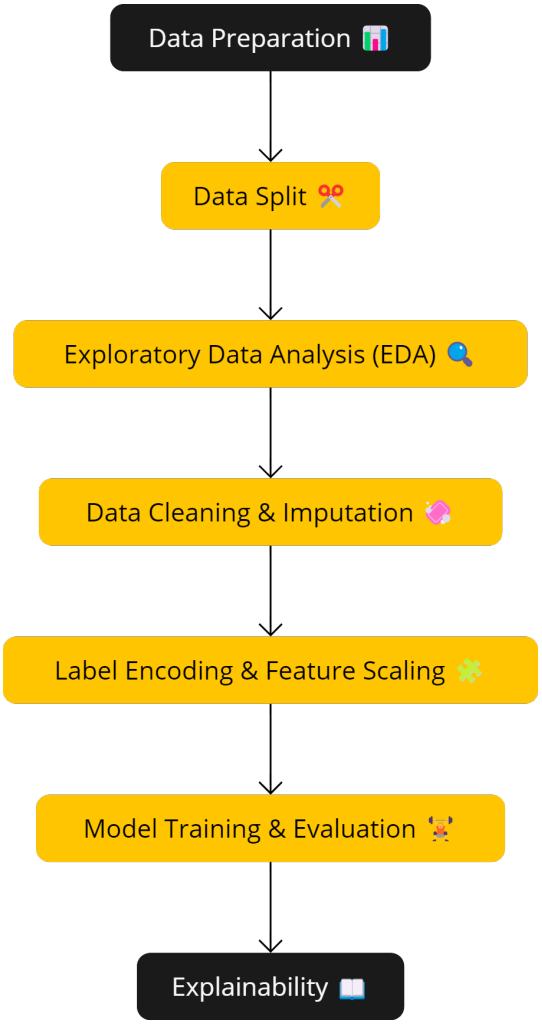


Figure 4.13: Process Flow

4.2.1 Splitting the Data into Training and Test Sets

To minimize overfitting and promote good generalization to new data, the first step is to divide the dataset. The training set is used to build and train the machine learning model, while the test set serves to assess how well the model performs on previously unseen data. This separation ensures that the model is trained on a substantial portion of the data but is evaluated independently on a different subset, allowing for an accurate assessment of its performance. Cross-validation techniques were also considered to improve the model's robustness by averaging results over different data subsets. The figue 4.1 and 4.2 represent the snapshot of the data used in training and testing.

4.2.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was performed to better understand the underlying structure of the dataset and to identify potential challenges, such as multicollinearity or outliers. Visual tools like histograms, scatter plots, and heatmaps were employed to analyze the distribution of key features and their correlation with energy usage. These insights guided the choice of models and feature engineering techniques, ensuring that the machine learning pipeline was built on solid foundations. Additionally, statistical tests were conducted to evaluate the normality and variance of the data, ensuring that the dataset was well-prepared for model training. This step has already been explained in the 4.1.2 section in detail.

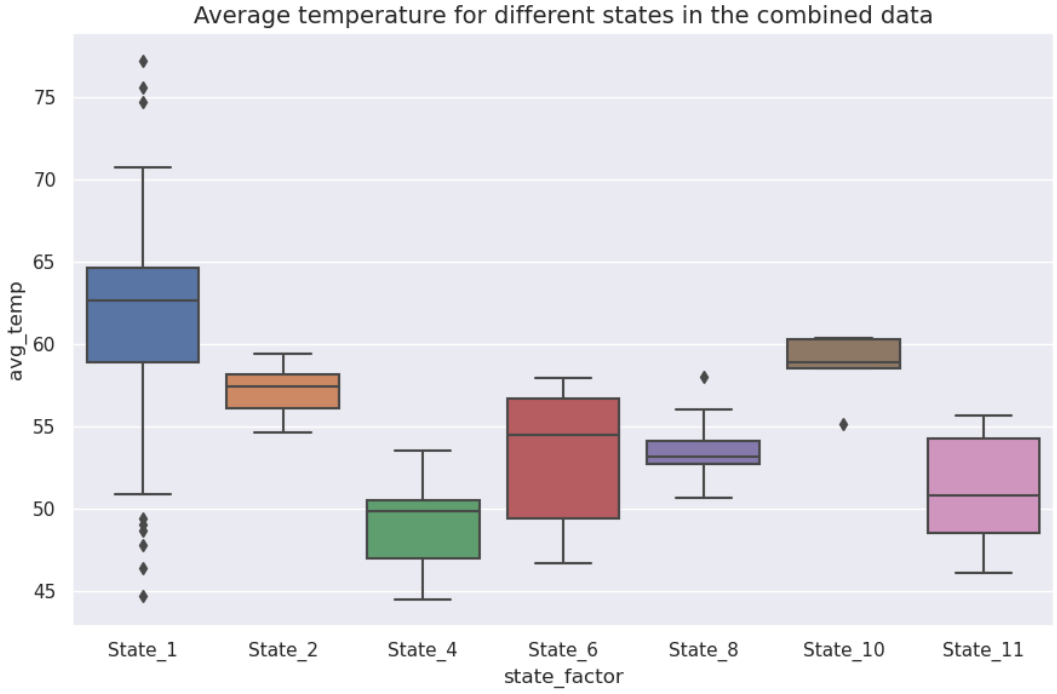


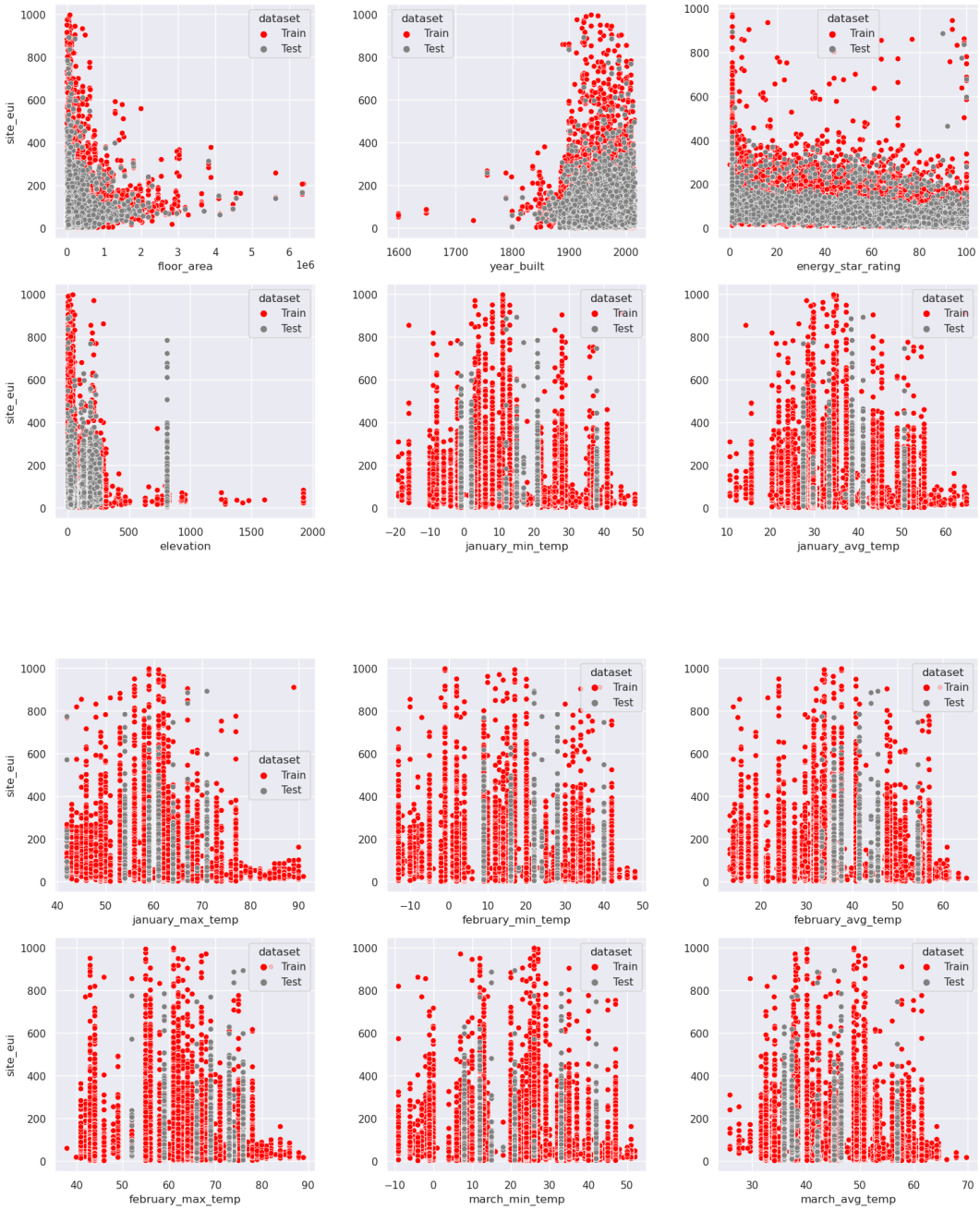
Figure 4.14: Temperature By State

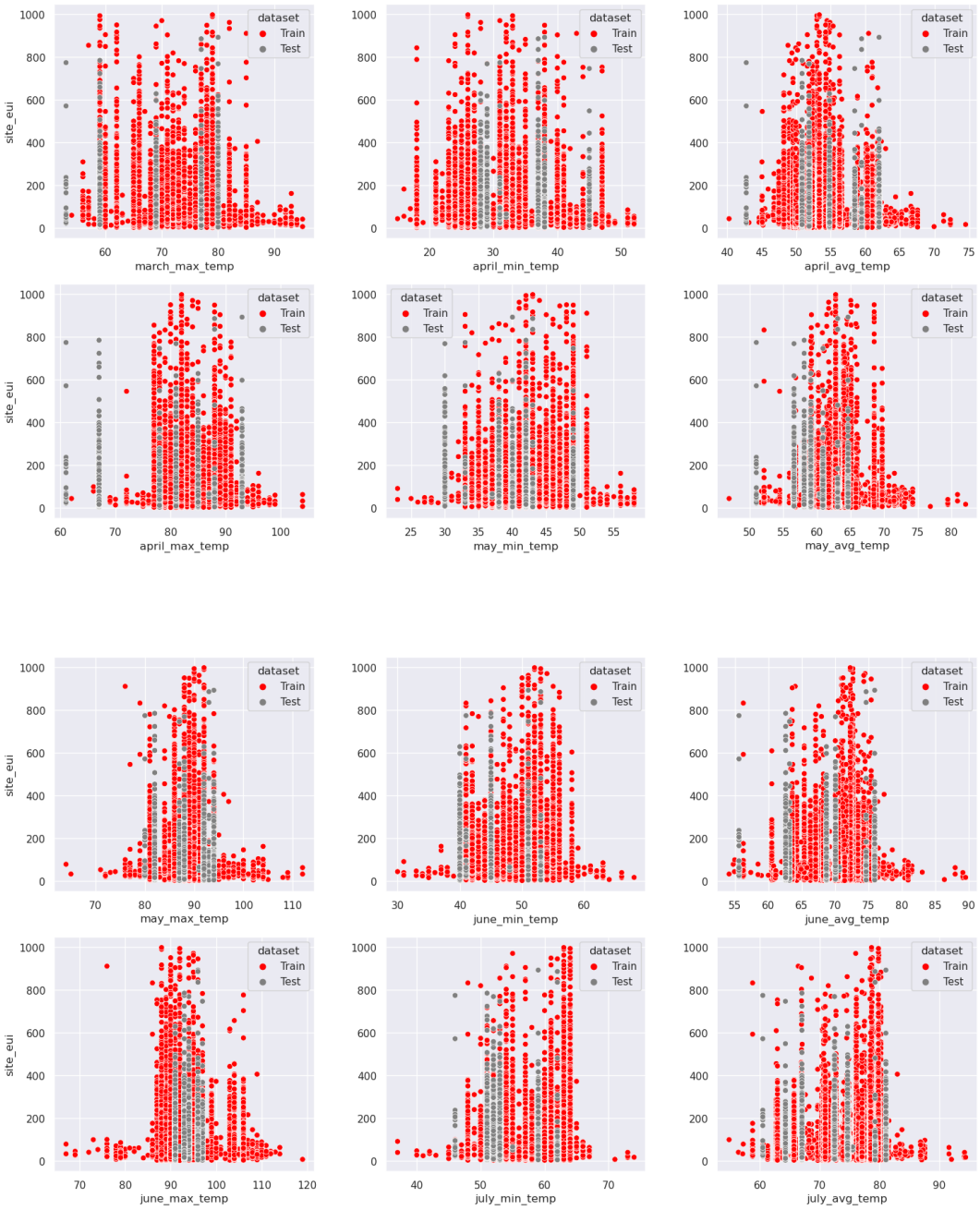
The next steps involved employing Boxplots to visually compare the distributions of average temperature across different states within the combined dataset. Among the states analyzed in the figure 4.14, State_1 exhibits the highest average temperature, while State_4 experiences the lowest. In terms of temperature variability, State_6, State_11, State_1, and State_4 demonstrate greater dispersion compared to State_2, State_8, and State_10.

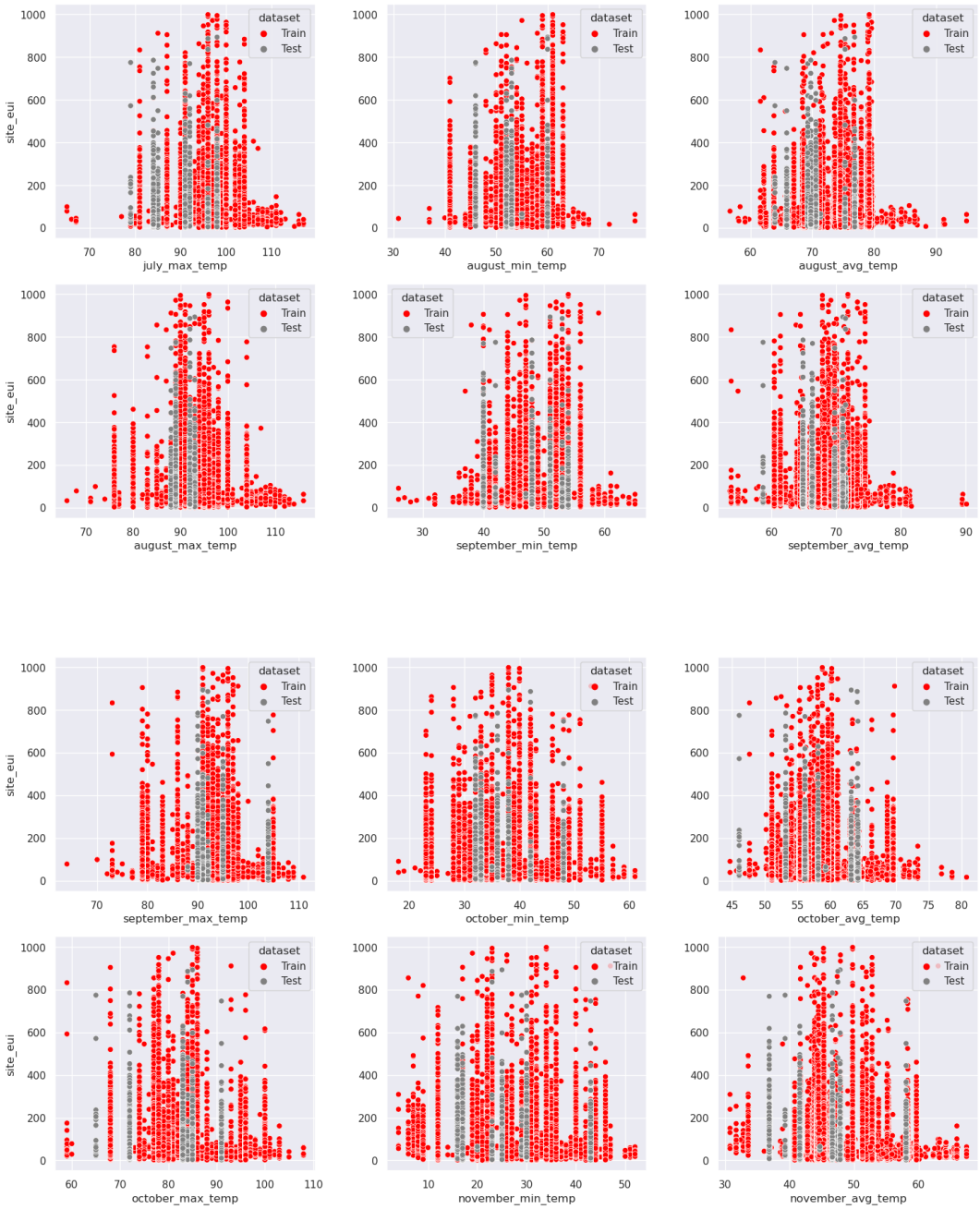
Correlation

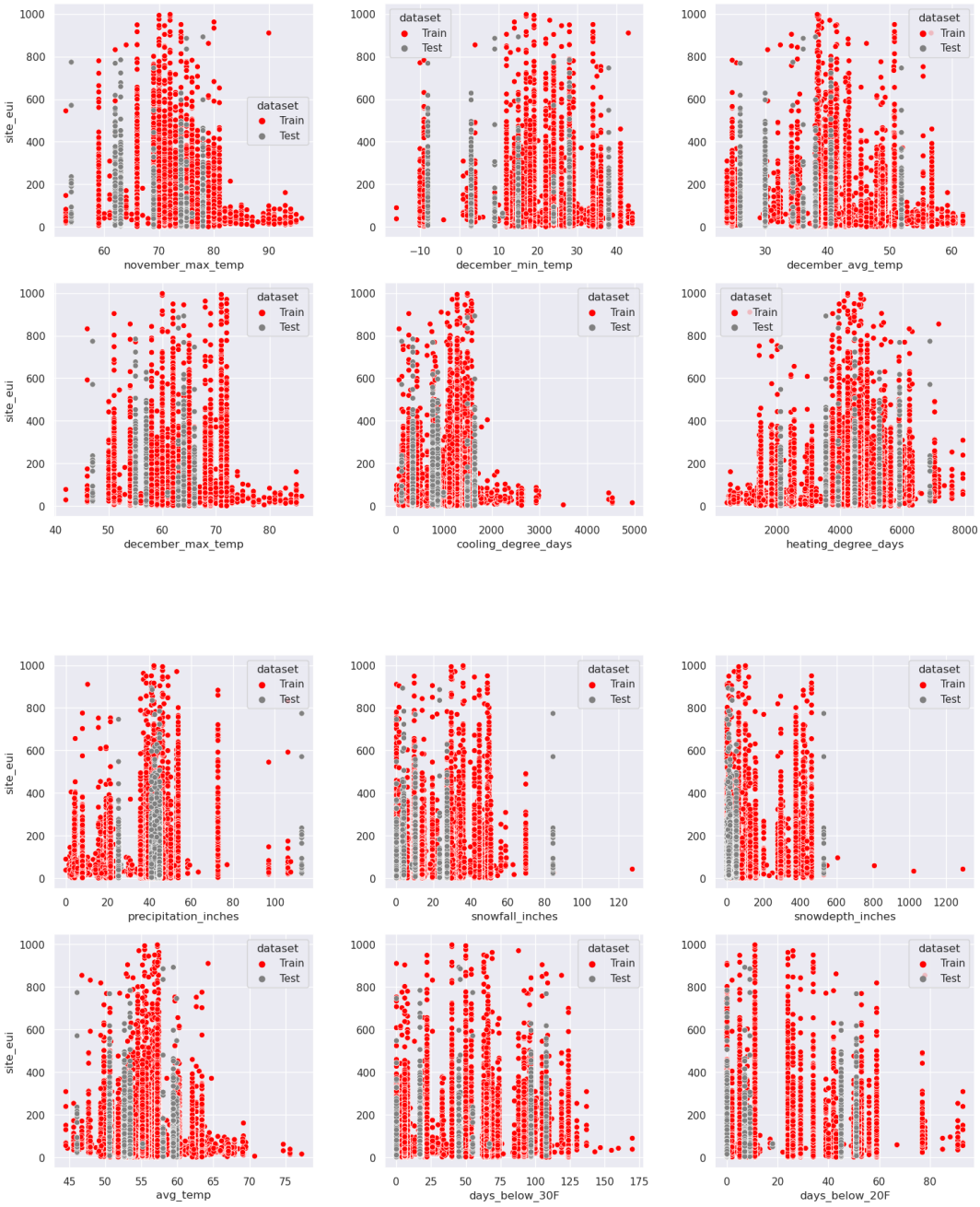
These images present a series of scatter plots comparing the train and test datasets for various temperature-related features and their relationship to the target variable,

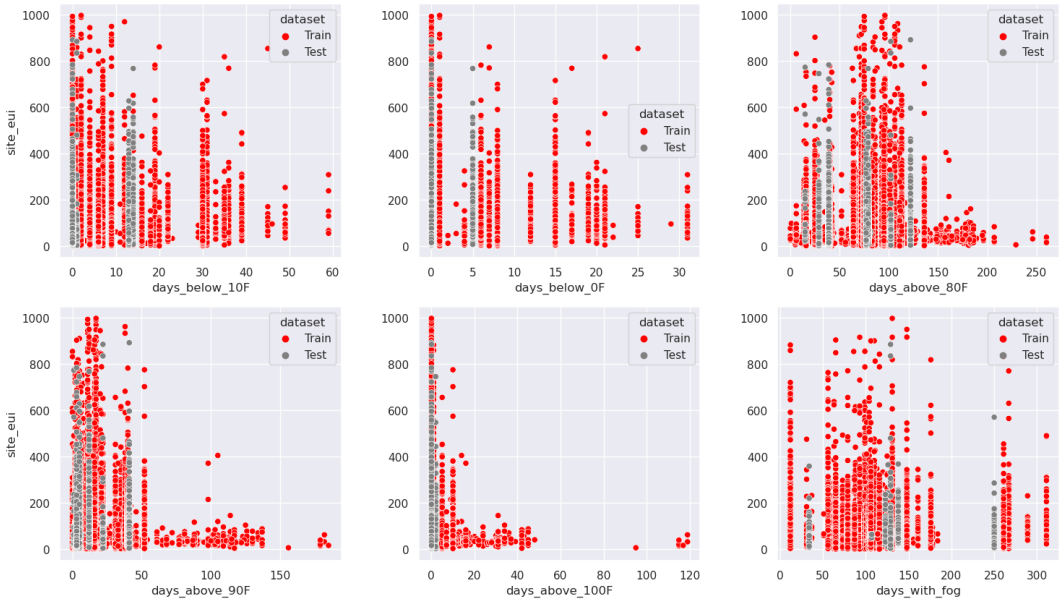
site energy usage. Each scatter plot shows red dots representing the training data and gray dots representing the test data. The x-axis of each plot represents different temperature-related features (e.g., minimum, maximum, average temperatures across different seasons), while the y-axis represents the site energy usage.




site_eui
dataset
Train
Test
march_max_temp
april_min_temp
april_avg_temp
april_max_temp
may_min_temp
may_avg_temp
may_max_temp
june_min_temp
june_avg_temp
june_max_temp
july_min_temp
july_avg_temp


site_eui
july_max_temp
august_min_temp
august_avg_temp
august_max_temp
september_min_temp
september_avg_temp
september_max_temp
october_min_temp
october_avg_temp
october_max_temp
november_min_temp
november_avg_temp
dataset
Train
Test


site_eui
dataset
Train
Test
november_max_temp
december_min_temp
december_avg_temp
december_max_temp
cooling_degree_days
heating_degree_days
precipitation_inches
snowfall_inches
snowdepth_inches
avg_temp
days_below_30F
days_below_20F



Summary

1. **Temperature Variability**: There does not seem to be a simple linear relationship between temperature features and energy usage. The scatter plots indicate that the relationship between temperature and energy consumption is likely non-linear or influenced by other factors (like building type, size, or insulation).

2. **Broad Energy Usage Range**: Energy usage is spread across a wide range for all temperature values in both training and test sets, suggesting that temperature alone is not a sufficient predictor of energy usage. This reinforces the need for machine learning models to consider a variety of features (e.g., facility type, energy star rating, floor area) in predicting energy consumption.

3. **Similar Distribution in Train and Test Sets**: Both the training and test sets appear to have similar distributions of energy usage across temperature ranges, which is important for ensuring that the model generalizes well across both datasets. However, the test set appears to have fewer samples than the training set, as indicated by the sparse gray dots, but the trend still follows the same pattern.

4. **Floor Area**: There is a **negative correlation** between floor area and energy usage intensity (EUI). Smaller buildings tend to have higher EUI, while larger buildings spread energy consumption over more space, leading to lower EUI.

5. **Year Built**: **Older buildings** generally have higher EUI, indicating lower energy efficiency. Newer buildings, built with modern standards and technologies, tend to be more energy-efficient with lower EUI.

6. **Energy Star Rating**: Buildings with **higher energy star ratings** are more energy-efficient, showing a **clear negative correlation** between energy star rating and EUI. Buildings with low ratings tend to consume more energy.

7. **Elevation**: Elevation does not show a strong influence on energy consumption across most of the dataset, with energy usage being similar at different elevations.

8. **January Minimum and Average Temperatures**: There is no strong linear correlation between January temperatures and EUI. Energy usage varies widely across different temperature ranges, suggesting other building characteristics (e.g., insulation, heating systems) have a larger impact on energy consumption than outdoor temperature alone.

4.2.3 Data Cleaning and Filling in Missing Values

This step is crucial for ensuring data quality. Missing values in critical features such as weather data and building characteristics were filled using imputation techniques. Simple imputation methods like replacing missing values with the median or mean were used where appropriate, while categorical variables were processed through mode imputation or one-hot encoding. The cleaning process also involved identifying and treating outliers that could skew the model's predictions. Outliers were handled either by capping their values or by excluding them, ensuring that the model learned from more reliable data.

During initial data exploration, several observations were made regarding missing values. The features days_with_fog, direction_peak_wind_speed, direction_max_wind_speed, and max_wind_speed contained a substantial amount of missing data, exceeding 50% in the training set and 88% in the test set. Due to this significant lack of information, these features were dropped from further analysis. In contrast, the feature energy_star_rating had 35.25% and 23.22% missing values in the training and test sets, respectively. The year_built feature exhibited a lower proportion of missing values, with 2.42% in the training set and 0.94% in the test set.

Despite the relatively low number of missing values in year_built, imputation was performed for both year_built and energy_star_rating. This decision stemmed from the observation that energy_star_rating showed the highest absolute correlation with the target variable. In the training set, energy_star_rating was missing for 26,709 observations, and year_built for 1,837 observations. There were 26,078 instances where exactly one of these features was missing, and 1,234 instances where both were missing, resulting in a total of 27,312 observations with missing data. In the test set, energy_star_rating was missing for 2,254 observations, and year_built

for 92 observations. Similarly, 2,262 observations had one missing feature, and 42 had both missing, leading to a total of 2,304 observations with missing data.

Since both energy_star_rating and year_built are numerical features, median imputation was employed to mitigate the potential impact of outliers. This imputation was first applied to the training set. Subsequently, the missing values in the test data for these two features were imputed using the medians calculated from the training data, ensuring consistency and preventing data leakage.

4.2.4 Label Encoding and Feature Scaling

It is used to prepare the data for machine learning algorithms. Label encoding converts categorical features to numerical features, and feature scaling transforms the features to a common scale. Since algorithms like Random Forest (RF), XGBoost, and CatBoost can handle numerical data, categorical features such as building type and state location were converted to numerical values using target encoding. Feature scaling was applied to ensure that all numerical features had a consistent range, which is particularly important for gradient boosting algorithms like XGBoost and CatBoost to converge efficiently.

The monthly weather statistics were compressed into seasonal weather statistics. This significantly reduced the number of features without substantial information loss. First, the monthly temperature statistics columns were separated into four lists corresponding to the four seasons. Next, seasonal temperature statistics were extracted from the monthly temperature statistics in the training DataFrame data_train. Additionally, **cooling_degree_days** and **heating_degree_days** were converted from a yearly scale to a monthly scale. This feature extraction procedure was then replicated for the test data. In total, 22 new features were generated from the original pool of 39 weather-related features.

Finally, the original weather-related features were dropped from both the training and test sets. The month-based temperature-related features were transformed into season-based features by partitioning the year into four seasons: winter (December, January, February), spring (March, April, May), summer (June, July, August), and autumn (September, October, November).

The following features were created from feature engineering:

- `min_temp_winter`: Minimum temperature in winter (in Fahrenheit) at the building's location.
- `max_temp_winter`: Maximum temperature in winter (in Fahrenheit) at the building's location.
- `avg_temp_winter`: Average temperature in winter (in Fahrenheit) at the building's location.
- `std_temp_winter`: Standard deviation of temperature in winter (in Fahrenheit) at the building's location.

- `skew_temp_winter`: Skewness of temperature in winter (in Fahrenheit) at the building's location.
- `cooling_degree_days_per_month`: Average number of degrees where the daily average temperature exceeds 65 degrees Fahrenheit in a month.
- `heating_degree_days_per_month`: Average number of degrees where the daily average temperature falls under 65 degrees Fahrenheit in a month.

Similarly, the features were defined for spring, summer, and autumn seasons.

4.2.5 Training and Hyperparameter Tuning of Machine Learning Algorithms

This is an iterative process. The model is trained on the training set and evaluated on the test set. The model's performance was improved by adjusting the hyperparameters of the algorithm, focusing on parameters like the number of trees, learning rate, and maximum depth in RF and XGBoost, as well as the number of iterations and learning rate in CatBoost.

In this section, three machine learning algorithms were employed — Random Forest, XGBoost, and CatBoost — to address the prediction task and compare their baseline performances.

Firstly, all numerical features are converted to the float64 data type. Subsequently, the predictor variables are separated from the target variable within both the training and test datasets. The building_id column is removed from both datasets as it is irrelevant to the prediction of the target variable.

Figure 4.15: Random Forest

Figure 4.16: XGBoost

```
# CatBoost
cat_feautres_idx = np.where(X_train.dtypes != 'float64')[0]
catb = CatBoostRegressor(iterations = 1000,
                         learning_rate = 0.02,
                         depth = 12,
                         eval_metric = 'RMSE',
                         # early_stopping_rounds = 42,
                         random_state = 0,
                         bagging_temperature = 0.2,
                         od_type = 'Iter',
                         metric_period = 100,
                         od_wait = 100)
catb.fit(X_train, y_train,
         eval_set = (X_test, y_test),
         cat_features = cat_feautres_idx,
         use_best_model = True,
         verbose = False)
```

[68] - Command executed in 6 min 43 sec 171 ms by Aman Singla on 12:30:00 PM, 8/21/24 PySpark (Python)

Figure 4.17: CatBoost

Hyperparameter Tuning

Hyperparameter tuning plays a critical role in optimizing machine learning models for energy usage prediction. While traditional methods such as grid search and random search are commonly employed, their exhaustive nature makes them inefficient for large, complex models. Optuna, a state-of-the-art hyperparameter optimization framework, offers a more flexible and efficient approach through its ability to perform adaptive search algorithms, such as Tree-structured Parzen Estimator (TPE) and CMA-ES (Covariance Matrix Adaptation Evolution Strategy) [4].

Optuna has proven to be particularly effective in optimizing ensemble models such as XGBoost, CatBoost, and Random Forests, which are often used in energy usage prediction tasks due to their ability to handle non-linear relationships and large datasets. By automating the search for optimal hyperparameters, Optuna allows models to achieve higher accuracy with fewer computational resources compared to traditional method [32].

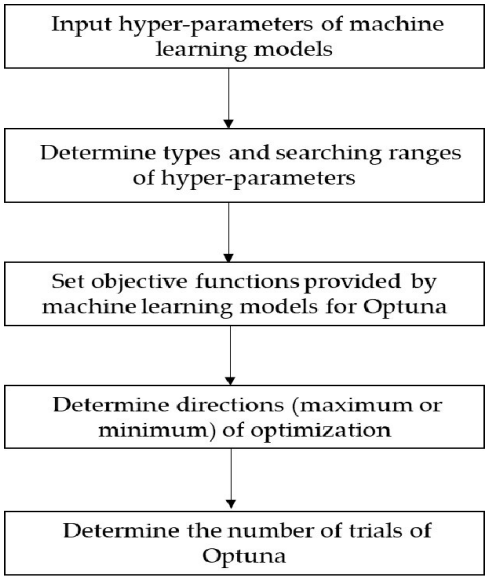


Figure 4.18: Optuna Process for Hyperparameter Tuning [35]

Given the superior performance of the Random Forest algorithm among the three baseline models, the Optuna framework was chosen to refine its hyperparameters. Specifically, the following hyperparameters were targeted for optimization:

- `n_estimators` (integer): The number of trees in the forest.
- `max_depth` (integer): The maximum depth of each tree.
- `min_samples_split` (integer): The minimum number of samples required to split an internal node.
- `max_features` (float): The proportion of features to consider when searching for the best split at each node.

```python
#Hyperparameter Tuning

# Objective function
def objective_rf(trial, data = X_train, target = y_train):
    param = {
        "n_estimators": trial.suggest_int("n_estimators", 100, 500),
        "max_depth": trial.suggest_int("max_depth", 5, 20),
        "min_samples_split": trial.suggest_int("min_samples_split", 2, 10),
        "max_features": trial.suggest_float("max_features", 0.01, 0.95),
        "random_state": 0
    }

    model = RandomForestRegressor(**param)
    kfolds = KFold(n_splits = 6, shuffle = True, random_state = 0)
    scores = cross_val_score(model, data, target, cv = kfolds, scoring = "neg_root_mean_squared_error")
    return -scores.mean()
```

✓ - Command executed in 237 ms by Aman Singla on 9:19:55 AM, 8/13/24 — Frozen — PySpark (Python)

Figure 4.19: Objective Function

```python
# Tuning function
def tuner(objective, n = 10, direction = 'minimize'):

    # Create Study object
    sampler = optuna.samplers.TPESampler(seed = 0)
    study = optuna.create_study(direction = direction, sampler = sampler)

    # Optimize the study
    study.optimize(objective, n_trials = n)
    display(optuna.visualization.plot_optimization_history(study))

    # Print the result
    best_params = study.best_params
    best_score = study.best_value
    print(f"Best RMSE score: {best_score}")
    print(f"Optimized parameters: {best_params}\n")
    print("-------- Tuning complete --------")

    # Return best parameters for the model
    return best_params, best_score
```

- Command executed in 217 ms by Aman Singla on 10:26:36 PM, 8/06/24 PySpark (Python)

Figure 4.20: Tuning Function

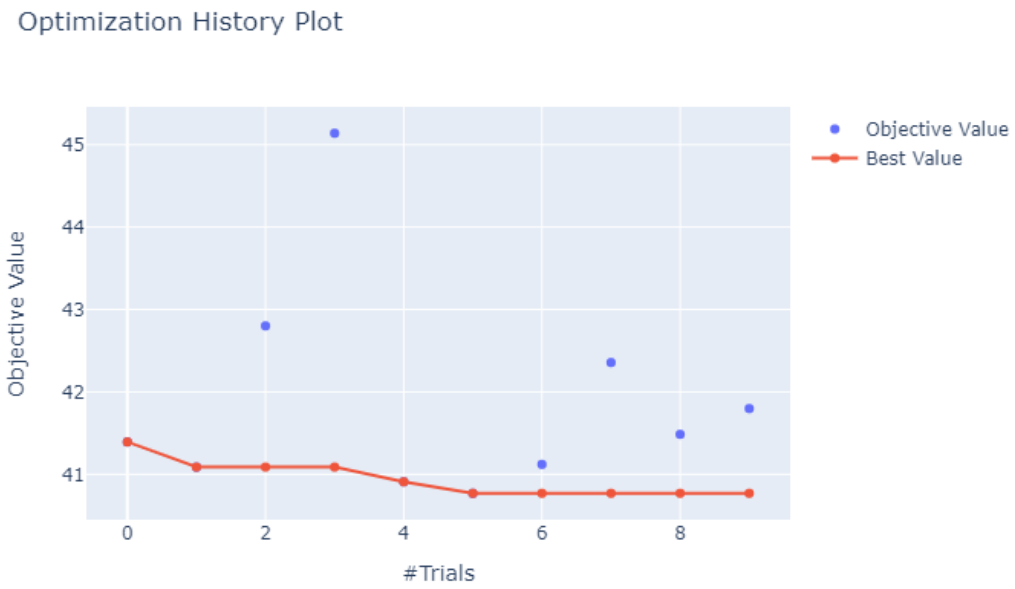


Figure 4.21: Optimized Plot

5 Results and Discussion

5.1 Explainability of AI

It is important for understanding why and how machine learning models make decisions. This can be done using a variety of methods, such as LIME and SHAP.

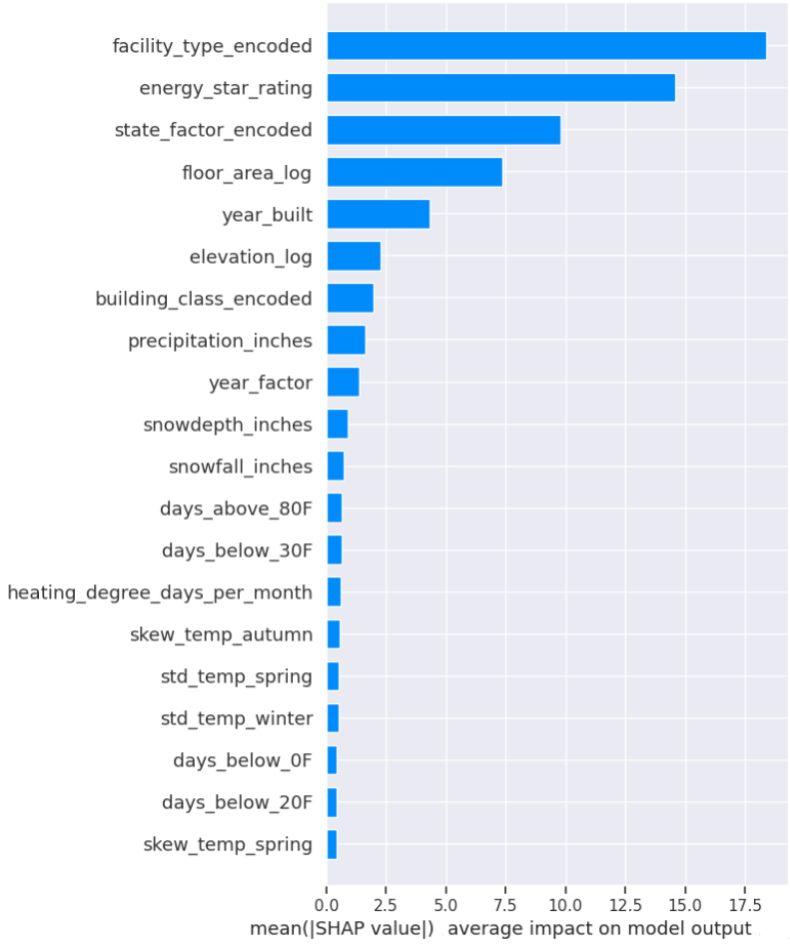


Figure 5.1: SHAP Variable Importance plot

The bar plot in the Figure 5.1 is a **SHAP Variable Importance plot**, which shows the **average impact** of each feature on the model's output based on their

mean(|SHAP value|). Essentially, the plot highlights the overall importance of each feature in the model, with the most important features listed at the top.

1. **Feature Importance**:
    - **facility_type_encoded** is the most important feature, with the highest mean SHAP value, meaning it has the largest average impact on the model's predictions.
    - **energy_star_rating** and **state_factor_encoded** also have significant impacts, following closely behind **facility_type_encoded** in terms of their importance.
    - Features like **floor_area_log** and **year_built** have moderate importance, while features like **days_below_30F** and **days_above_80F** have relatively low importance.
2. **Magnitude of Impact**:
    - The x-axis represents the mean absolute SHAP value, which quantifies how much, on average, each feature contributes to changing the model's predictions.
    - Larger values indicate features that more frequently shift the model's predictions by a significant amount. For example, **facility_type_encoded** has the highest average impact, influencing the predictions more strongly than other features.
    - In contrast, features with lower values like **skew_temp_spring** and **days_below_20F** have minimal influence on the model's output.
3. **Order of Importance**:
    - The features are ordered from top to bottom based on their average SHAP impact. This allows for a clear understanding of which features the model relies on most when making predictions.
    - Features related to the **type of facility**, **energy star rating**, and **location (state_factor_encoded)** dominate, indicating that the model places a lot of weight on these attributes when predicting energy usage.

Summary: Facility type, energy star rating, and state factor are the top three most important features in determining the model's predictions. These variables have the greatest average influence on the output. Less impactful features, such as temperature skewness and days below freezing, contribute minimally to the model, indicating that they don't significantly affect the predicted energy usage in this particular model.

This SHAP bar plot provides a clear ranking of feature importance, helping to understand which variables the model prioritizes most in energy usage prediction.

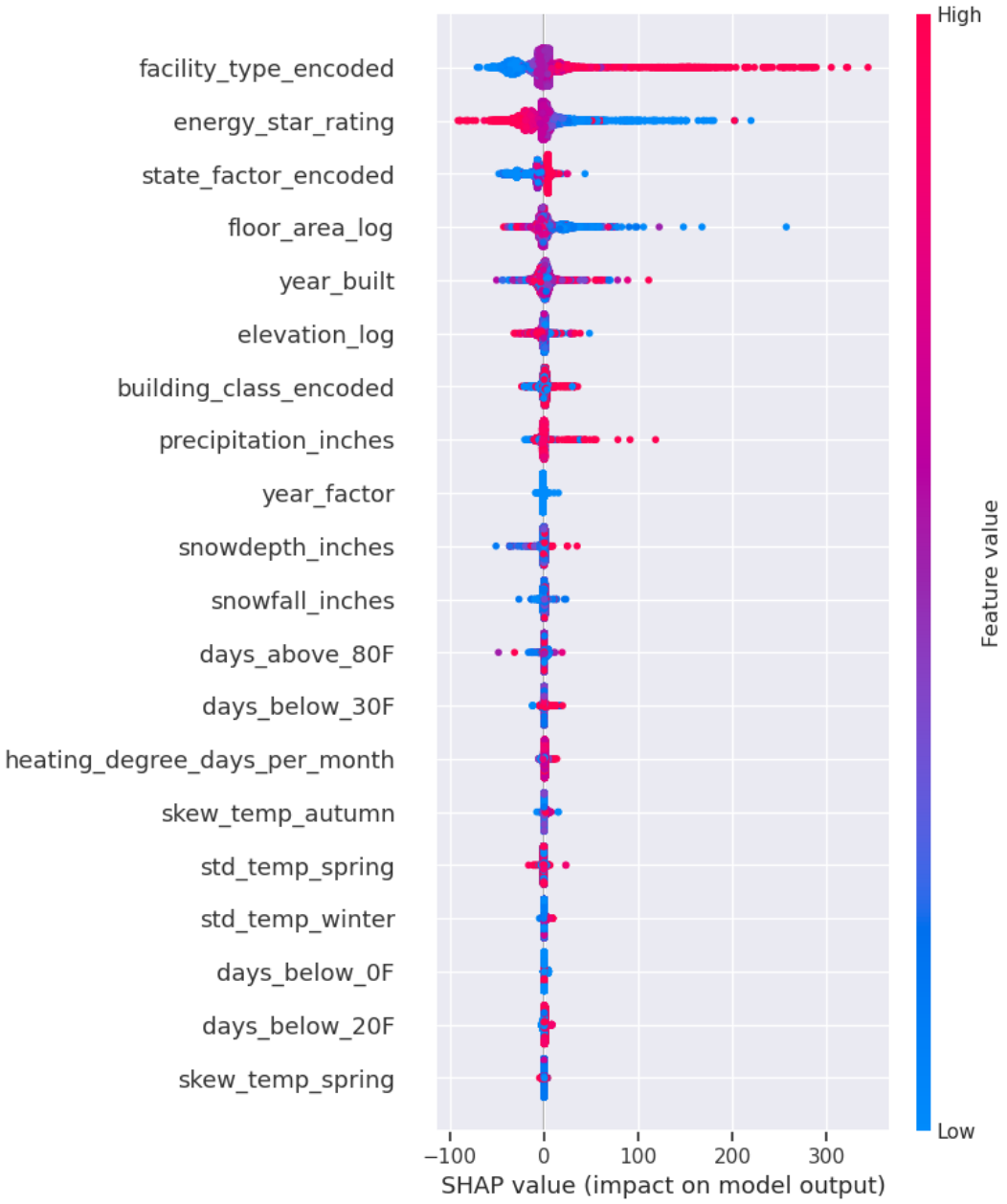


Figure 5.2: SHAP Summary plot

The plot represented in Figure 5.2 is a SHAP summary plot, which visualizes the impact of various features on the model's predictions using SHAP (SHapley Additive exPlanations) values. Each point on the plot represents a single prediction, with the position along the x-axis indicating the SHAP value (impact on the model's output), and the color representing the feature value (from low in blue to high in

red).

1. **Feature Importance**:
    - Features are ranked in order of their importance, with the most impactful feature at the top. In this case, **facility_type_encoded** has the largest influence on the model's output, followed by **energy_star_rating** and **state_factor_encoded**. The least important features are at the bottom, such as **skew_temp_spring** and **days_below_20F**.
2. **SHAP Values (Impact on Model Output)**:
    - The x-axis represents the SHAP values, which measure how much a feature contributes to increasing or decreasing the prediction.
    - Positive SHAP values (to the right) increase the model's prediction, while negative SHAP values (to the left) decrease it.
3. **Feature Values (Color Encoding)**:
    - The color represents the value of each feature for a specific data point.
        - **Red** indicates high feature values, while **blue** indicates low values.
    - For example, in the case of **facility_type_encoded**, higher values (red) generally push the SHAP value to the right, increasing the prediction, while lower values (blue) push it to the left, decreasing the prediction.
4. **Feature Behavior**:
    - For some features, such as **energy_star_rating**, high values (red) tend to have a negative SHAP value, meaning they reduce the predicted output. In contrast, high values of **floor_area_log** and **facility_type_encoded** tend to have positive SHAP values, increasing the model's predictions.

Summary: Facility type, energy star rating, and state factor are the most influential features in the model's predictions. The SHAP plot provides a clear visualization of how these features influence the model, with color encoding allowing the user to see how high or low feature values affect the output. Overall, features like facility_type_encoded and floor_area_log are driving preditions higher, while higher values of energy_star_rating tend to lower the model's predictions.

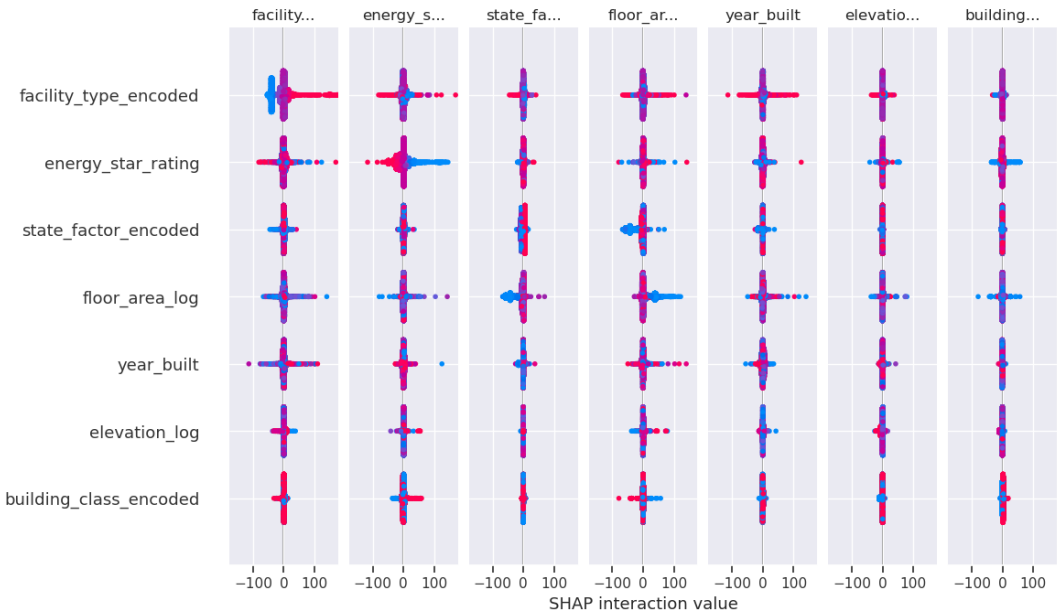


Figure 5.3: SHAP Interaction Plot

The plot shown in Figure 5.3 is a **SHAP interaction values plot**, which shows how pairs of features interact to influence the model's predictions. Each column represents one feature and each row represents another feature. The x-axis shows the **SHAP interaction values**, which indicate how much the interaction between the two features contributes to the model's prediction.

1. **Feature Interaction**:
    - The plot visualizes the interaction effects between pairs of features. Each point represents a SHAP interaction value for a specific observation.
    - The x-axis indicates the SHAP interaction value for that feature pair, with values closer to zero meaning little interaction effect, and values far from zero (positive or negative) showing significant interaction effects.
2. **Color Mapping**:
    - Similar to other SHAP plots, the color of each point represents the value of the feature being evaluated, ranging from **low (blue)** to **high (red)**.
    - For instance, in the top-left box, **facility_type_encoded** interacts with itself (as seen by the clear spread), and the color reveals how different values of the feature influence these interactions.
3. **Diagonal Elements**:
    - The diagonal elements (e.g., the top-left box, the second box in the second column, etc.) represent a feature interacting with itself. This helps to

visualize the feature's own contribution to the model output without any other feature's influence.

- These diagonal boxes show the standard SHAP value spread (similar to the summary plot), where red indicates higher feature values that have a positive or negative influence on the prediction, and blue shows the effect of lower feature values.

4. **Off-Diagonal Interactions**:
    - Off-diagonal elements represent interactions between two different features. For instance, the interaction between **facility_type_encoded** and **energy_star_rating** (first row, second column) or between **year_built** and **floor_area_log** (fifth row, fourth column).
    - A wide spread in the SHAP interaction values indicates a significant interaction between those two features. For example, if the points are dispersed both positively and negatively on the x-axis, it suggests that the interaction between the two features has a strong impact on model predictions.

Summary:
Facility type encoded seems to have strong interaction effects with features like energy star rating, state factor encoded, and floor area log, indicating that the type of facility impacts how these other features influence energy usage predictions. Energy star rating has some interaction effects with state factor encoded and year built, showing that the model adjusts its predictions based on the combination of energy efficiency rating and the state or the building's age. Interactions between top features like facility type encoded, energy star rating, and floor area log indicate that these combinations significantly impact the energy usage predictions. For instance, the type of facility influences how much the energy star rating or the building's size matters for the model's output. This SHAP interaction plot provides deeper insights beyond individual feature importance by showing how features work together to drive predictions.

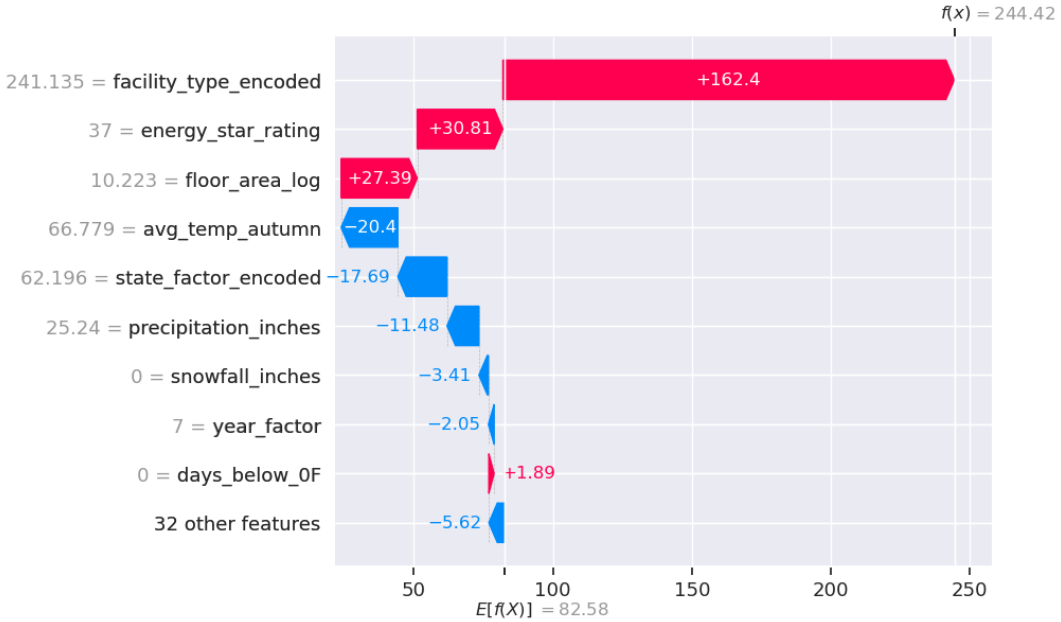


Figure 5.4: SHAP Waterfall Plot

The **SHAP waterfall plot** in Figure 5.4 breaks down the contributions of individual features to a specific prediction by showing how each feature impacts the model's output relative to the baseline.

Key Elements of the SHAP Waterfall Plot:

1. **E[f(x)] = 82.58**: This is the **baseline value**, which represents the average prediction of the model across all observations in the dataset. The baseline is the starting point from which the individual prediction deviates.

2. **f(x) = 244.42**: This is the **final prediction** for the specific instance. The features either contribute to increasing or decreasing this prediction relative to the baseline.

3. **SHAP values**: Features contribute positively (shown in **red**) or negatively (shown in **blue**) to push the prediction away from the baseline. The length of the bars indicates the magnitude of each feature's impact on the final prediction.

Explanation of the Features' Contributions:

- **facility_type_encoded**: This feature has the **largest positive contribution** (+162.4), which pushes the prediction significantly upward. It suggests that the specific facility type of this observation is strongly associated with higher energy consumption.

- **energy_star_rating**: Contributing **+30.81**, this feature also pushes the prediction upward. Even though higher energy star ratings typically indicate energy efficiency, in this particular case, the rating is contributing to higher energy usage for this instance.
- **floor_area_log**: This adds **+27.39** to the prediction. Larger buildings (represented by a higher floor area) tend to consume more energy, which is reflected in this positive contribution.
- **avg_temp_autumn**: This feature has a **negative impact** of **-20.4** units, meaning that higher average temperatures in autumn lead to reduced energy consumption, possibly because less heating is required.
- **state_factor_encoded**: This feature contributes **-17.69**, pulling the prediction down. The state in which the building is located likely has regulations, climate, or consumption patterns that reduce energy usage.
- Other smaller negative contributions include **precipitation_inches** (-11.48), **snowfall_inches** (-3.41), and **year_factor** (-2.05), which slightly decrease the overall prediction.

Main Findings from the Waterfall Plot:

1. **Facility type** is by far the most influential feature for this specific prediction, strongly increasing energy consumption. The model associates this type of facility with significantly higher energy usage.
2. **Energy Star Rating** and **Floor Area** are also major contributors to the increase in the prediction, suggesting that for this particular instance, a larger building with a higher energy star rating still consumes a lot of energy.
3. **Average Temperature in Autumn** and **State Factor** contribute to lowering the prediction, indicating that warmer temperatures in autumn reduce energy needs, and the location of the building likely influences energy policies or habits that decrease energy usage.
4. Overall, the final prediction of **244.42** is substantially higher than the baseline value due to the strong positive influence of factors like facility type, energy star rating, and floor area.

This waterfall plot helps visually explain how each feature contributes to the deviation from the baseline prediction for this specific instance, with the largest impact coming from the facility type.

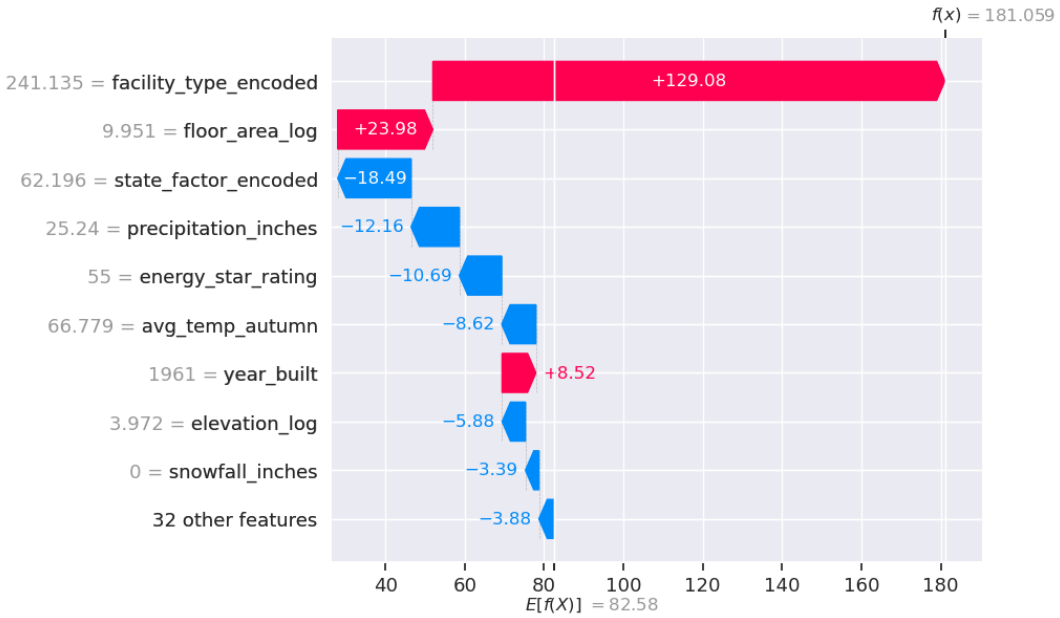


Figure 5.5: SHAP Waterfall Plot

The **SHAP waterfall plot** in Figure 5.5 represents another instance of the same model output set.

Explanation of the Features' Contributions:

1. **facility_type_encoded**: This feature has the **largest positive impact**, contributing **+129.08** to the prediction. This suggests that the type of facility in this instance is strongly associated with higher energy consumption.

2. **floor_area_log**: Contributes **+23.98**, indicating that the larger floor area of the building is significantly increasing the predicted energy usage.

3. **state_factor_encoded**: This feature has a **negative impact** of **-18.49**, pulling the prediction down. This likely reflects the influence of the building's location (state), where certain factors like regulations, climate, or energy habits reduce energy usage.

4. **precipitation_inches**: Contributes **-12.16**, suggesting that more precipitation correlates with lower energy consumption, possibly because cooler or wetter conditions reduce energy needs like cooling.

5. **energy_star_rating**: Reduces the prediction by **-10.69**, meaning that a higher energy star rating is contributing to more energy-efficient outcomes, thus lowering the predicted consumption.

6. **avg_temp_autumn**: Reduces the prediction by **-8.62**, likely due to warmer autumn temperatures decreasing energy needs for heating.

7. **year_built**: Contributes **+8.52**, meaning that the building's construction year (1961 in this case) is associated with higher energy consumption, possibly due to older building infrastructure or less efficient energy standards.

8. **elevation_log**, **snowfall_inches**, and other smaller features contribute negative SHAP values, slightly decreasing the prediction.

Main Findings from the Waterfall Plot:

- The final prediction of **181.059** is significantly higher than the baseline value of **82.58**, largely driven by the **facility type** and **floor area**. These two features alone account for the majority of the upward adjustment in the prediction.
- **State factor**, **precipitation**, and **energy star rating** play important roles in lowering the prediction, reflecting influences such as location, climate conditions, and energy efficiency standards.
- Overall, the **facility type** has the strongest influence on the prediction, driving the energy usage significantly higher, while other factors like **state**, **precipitation**, and **energy star rating** help to moderate this by reducing the predicted consumption.

The figure 5.6 represents a **SHAP decision plot**, which visualizes how different features contribute cumulatively to the model's final prediction for multiple instances. Each line represents a single observation, tracing the cumulative impact of features as the model progresses through them, starting from a baseline prediction and ending at the final output value.

Explanation of the SHAP Decision Plot:

1. **Model Output Value (x-axis)**:
    - The x-axis represents the model's output value which is predicted energy consumption. Values to the right of zero indicate higher predictions, while values to the left represent lower predictions.
2. **Features (y-axis)**:
    - The y-axis lists the features in descending order of their importance (top features are more influential).
    - Each feature adjusts the prediction cumulatively. As the line moves from top to bottom, it shows how each feature pushes the prediction higher (to the right) or lower (to the left).
3. **Lines**:
    - Each line corresponds to a single instance, tracing the step-by-step contribution of each feature to the final prediction.

- If a line moves to the right after a feature, that feature increases the prediction for that instance. If the line moves left, the feature decreases the prediction.
- The lines are color-coded from **blue to red** based on the magnitude of the final prediction (with **blue** for low values and **red** for high values).

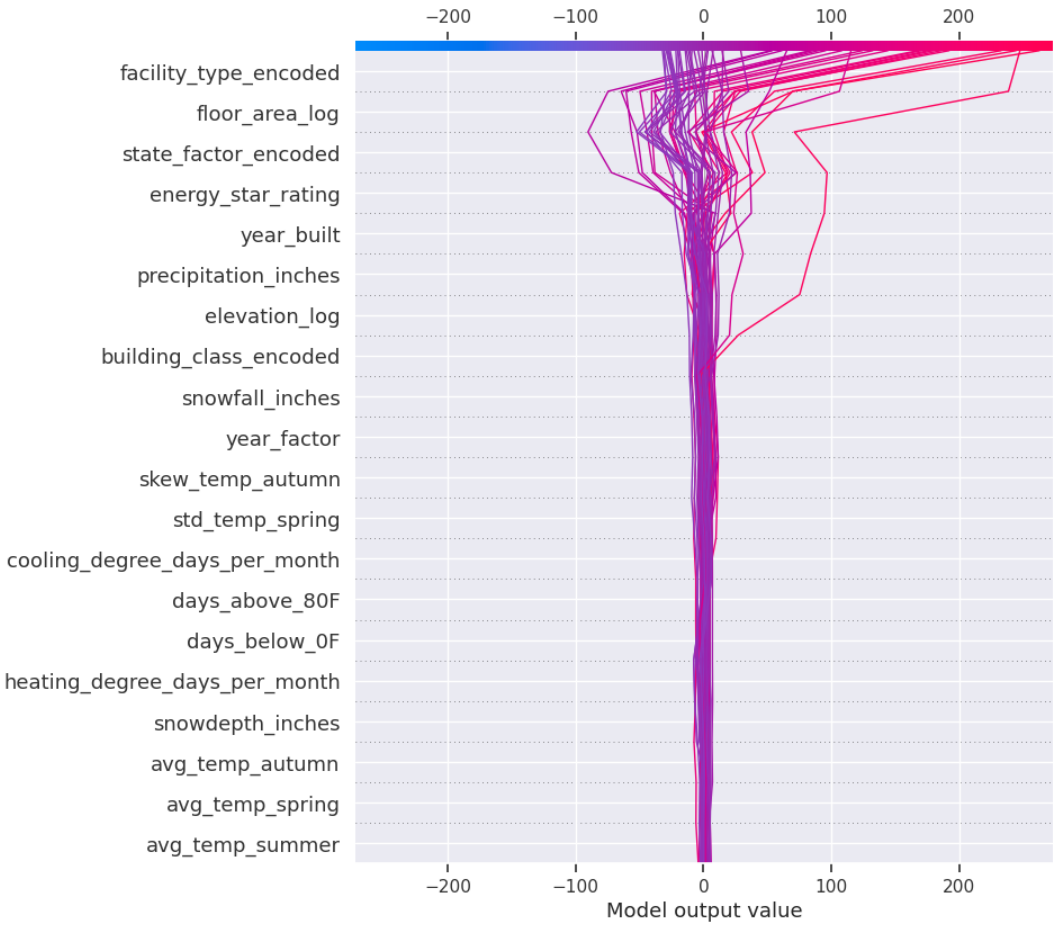


Figure 5.6: SHAP Decision Plot

Feature Contributions:

- **facility_type_encoded**: This feature has the largest impact, with many lines shifting substantially to the right after passing through it. This shows that the **facility type** increases the predicted energy usage for most instances, with some predictions going as high as 200+.
- **floor_area_log**: The feature **floor area** also increases the prediction for many instances, contributing to larger final outputs. Buildings with larger floor areas

generally have higher energy consumption, and this trend is visible as the lines shift further right for higher values of this feature.

- **state_factor_encoded** and **energy_star_rating**: **State factor** and **energy star rating** have more mixed effects. Some lines move to the left (reducing the prediction), while others move to the right. This indicates that these features contribute positively to some predictions and negatively to others, reflecting variation in how these factors influence energy consumption.
- **Precipitation and Temperature-Related Features**: Features like **precipitation_inches**, **avg_temp_autumn**, and **snowfall_inches** tend to pull many of the predictions leftward (lower), reflecting a reduction in energy usage for certain buildings in climates with more precipitation or higher autumn temperatures.

Summary:
The **facility_type_encoded** is the most influential feature, consistently driving higher predictions for energy usage across instances. This indicates that certain facility types are highly energy-consuming. The **floor_area_log** also significantly impacts the model's output, as larger buildings are expected to consume more energy. The **state_factor_encoded** and **energy_star_rating** have mixed effects, sometimes increasing and sometimes reducing energy predictions, depending on the specific instance. Climate-related features like **precipitation_inches** and **temperature** generally decrease the prediction, suggesting that buildings in cooler or wetter climates may consume less energy. Overall, the trend shows that larger facilities and certain facility types lead to much higher energy usage predictions, while geographic and climate-related factors can reduce the predicted energy consumption for certain buildings. This plot provides a clear visual breakdown of how each feature cumulatively impacts the model's predictions across multiple instances, highlighting key drivers of energy consumption.

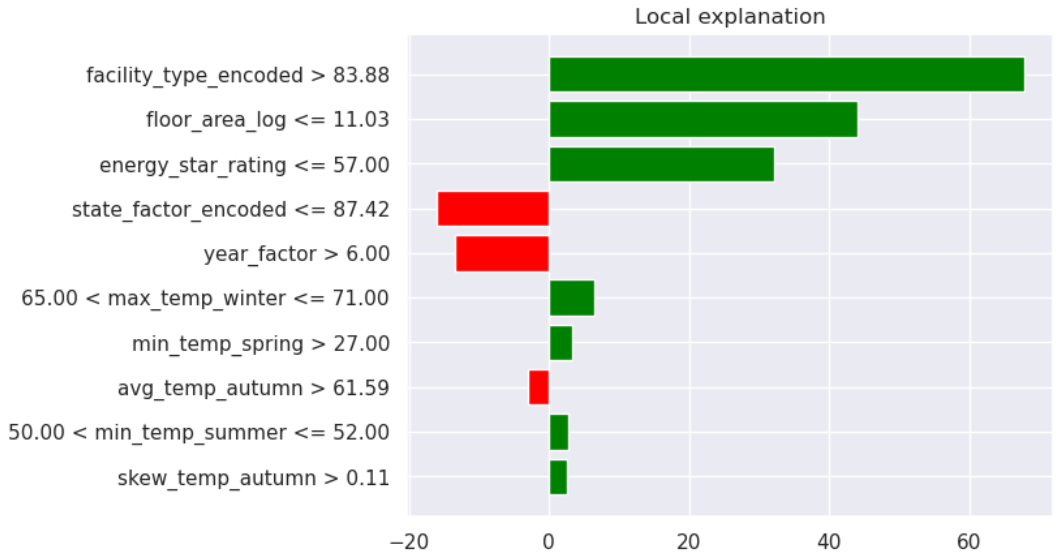


Figure 5.7: LIME Local Interpretation

The figure 5.7 represents a **LIME (Local Interpretable Model-agnostic Explanations) chart**, which provides a **local explanation** for a specific prediction made by the machine learning model. The plot visualizes the contributions of individual features to the final prediction for one instance.

Key Information:

- **Intercept**: 68.33 is the baseline prediction (the prediction the model would make if no features were considered).
- **Prediction_local**: 194.55 is the model's prediction based on the local explanation (LIME approximation).
- **Right (Actual Prediction)**: 181.06 is the model's actual prediction for this instance.

Explanation of Feature Contributions:

1. **facility_type_encoded > 83.88**: This feature has the largest positive contribution, adding around **+65** to the prediction. The facility type is a significant driver of higher energy consumption in this instance.
2. **floor_area_log <= 11.03**: This feature contributes about **+50** to the prediction. A smaller floor area (as indicated by the logarithmic transformation) increases the predicted energy consumption, possibly due to high energy use per square foot.

3. **energy_star_rating <= 57.00**: This feature contributes **+40** to the prediction. A lower energy star rating (less than 57) is associated with higher energy usage, reflecting the building's lower energy efficiency.

4. **state_factor_encoded <= 87.42**: This feature contributes **+30** to the prediction, indicating that the state in which the building is located tends to have higher energy consumption.

5. **year_factor > 6.00**: This feature reduces the prediction by **-15**, suggesting that buildings with a higher year factor (newer or renovated buildings) tend to have lower energy consumption, potentially due to better energy standards.

6. **65.00 < max_temp_winter <= 71.00**: This feature adds a small **+8** to the prediction. Warmer winter temperatures slightly increase energy consumption in this case, likely due to reduced heating needs but higher cooling demands.

7. **min_temp_spring > 27.00**: This feature has a minor positive contribution (**+5**), indicating that higher minimum spring temperatures correlate with slightly higher energy usage.

8. **avg_temp_autumn > 61.59**: This feature has a small **negative contribution** of **-3**, reducing the prediction slightly. Warmer autumn temperatures reduce energy consumption in this instance, possibly due to less heating required.

9. **Other minor features**: Several smaller features, such as **min_temp_summer** and **skew_temp_autumn**, have minor contributions to the prediction, with both positive and negative impacts, though these are relatively small compared to the top features.

Summary:

The **facility type** is the most significant factor contributing to the increase in energy usage for this instance, followed closely by the **floor area** and **energy star rating**. **State factor** and **year factor** also play notable roles, with the state pushing the prediction higher and the year factor slightly reducing it. Temperature-related features (such as winter and autumn temperatures) have smaller, yet relevant impacts on energy usage, with warmer temperatures either slightly increasing or reducing the final prediction.

Overall, the combination of these factors results in the **local prediction of 194.55** using LIME, which is slightly higher than the actual model prediction of **181.06**.

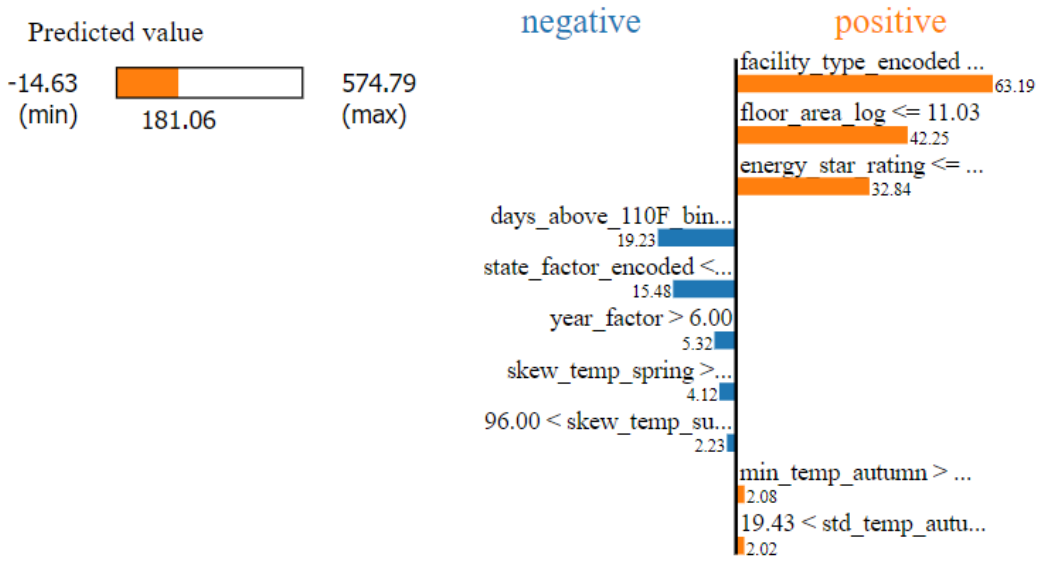


Figure 5.8: LIME notebook

Feature	Value
facility_type_encoded	241.14
floor_area_log	9.95
energy_star_rating	55.00
days_above_110F_binary	0.00
state_factor_encoded	62.20
year_factor	7.00
skew_temp_spring	0.42
skew_temp_summer	97.00
min_temp_autumn	43.00
std_temp_autumn	19.84

Figure 5.9: LIME notebook

The figures 5.8 and 5.9 represent **LIME** plot in notebook format and provides an explanation for an instance of model prediction. It breaks down the impact of individual features on the final prediction and compares it to the baseline.

Key Information:

1. **Intercept**: 90.39 is the baseline prediction (i.e., the value the model would predict if no features were considered).
2. **Prediction_local**: 186.38 is the prediction based on the local explanation provided by LIME. This is a slightly higher approximation of the model's actual prediction.
3. **Right (Actual Prediction)**: The actual model prediction is 181.06.

Explanation of Feature Contributions:
Positive Contributions (Pushing the Prediction Higher):

1. **facility_type_encoded > 83.88**: Contributes **+63.19**, significantly increasing the predicted energy usage. The facility type is the most influential factor in driving the energy consumption prediction upwards.
2. **floor_area_log <= 11.03**: Adds **+42.25** to the prediction. A larger floor area, as indicated by the logarithmic value, is associated with increased energy consumption.
3. **energy_star_rating <= 55.00**: Contributes **+32.84** to the prediction. A lower energy star rating, indicating less energy efficiency, drives the energy consumption prediction higher.
4. **min_temp_autumn = 43.00**: Has a smaller contribution of **+2.08**, indicating that lower autumn temperatures contribute slightly to higher energy usage, possibly due to increased heating needs.
5. **std_temp_autumn = 19.84**: Adds a minor contribution of **+2.02**, indicating that variation in autumn temperatures slightly increases energy consumption.

Negative Contributions (Pushing the Prediction Lower):

1. **days_above_110F_binary = 0.00**: Reduces the prediction by **-19.23**. The absence of days above 110°F indicates lower cooling needs, which reduces energy usage.
2. **state_factor_encoded <= 62.20**: Contributes **-15.48** to reducing the prediction. The state in which the building is located likely has characteristics (such as favorable regulations or climate) that reduce energy usage.
3. **year_factor > 6.00**: Decreases the prediction by **-5.32**, indicating that newer buildings or renovations (as captured by this factor) are associated with lower energy consumption due to better energy efficiency standards.

4. **skew_temp_spring > 0.42**: Has a small negative contribution of **-4.97**, indicating that certain temperature patterns in spring slightly reduce energy usage.

5. **skew_temp_summer = 97.00**: Contributes **-2.23**, further lowering the energy prediction. This could reflect patterns in summer temperatures that reduce energy usage.

Summary:
facility_type_encoded, **floor_area_log**, and **energy_star_rating** are the strongest contributors pushing the prediction higher, with the facility type having the greatest influence. **days_above_110F_binary** and **state_factor_encoded** provide significant negative contributions, reducing the energy prediction due to the absence of extreme temperatures and the state's characteristics that lead to reduced energy usage. Other climate-related factors, such as autumn and spring temperatures, play minor roles, with **min_temp_autumn** slightly increasing the prediction and **skew_temp_spring** reducing it. The **LIME local explanation** provides a prediction of **186.38**, which is close to the actual model prediction of **181.06**, showing that the most important factors contributing to the energy usage prediction are facility type, building size, energy efficiency, and state location.

5.2 Evaluation Results

5.2.1 ML Evaluation

By employing the diverse metrics mentioned in ML Evaluation Metrics section (3.3.1), a holistic evaluation of model performance was achieved, enabling a comparative analysis and facilitating informed decision-making regarding model selection and deployment. Tables 5.1, 5.2 and 5.3 represent the performances by Random Forest, XGBoost and CatBoost algorithms respectively.

Metric	Train	Test	CV_mean	CV_sd
RMSE	14.5282	28.1795	39.9145	0.6494
MAE	7.2849	15.5066	20.0986	0.2545
R^2	0.9378	0.7804	0.5303	0.0128

Table 5.1: Random Forest Performance Metrics

Metric	Train	Test	CV_mean	CV_sd
RMSE	33.038887	32.623326	40.340473	0.724725
MAE	18.549213	18.630172	20.834601	0.250881
R^2	0.678350	0.705648	0.520303	0.010552

Table 5.2: XGBoost Performance Metrics

Metric	Train	Test	CV_mean	CV_sd
RMSE	35.389342	35.595565	41.552393	0.619751
MAE	19.301683	20.456707	21.127464	0.213137
R^2	0.630956	0.649569	0.490991	0.011497

Table 5.3: CatBoost Performance Metrics

Figures 5.10 and 5.11 represent the evaluation metrics (RMSE, MAE, and R^2) for cross validation and test scores respectively. The Random Forest regressor demonstrated the best overall performance for energy usage prediction. It had the lowest error rates (RMSE and MAE) and explained the highest proportion of variance in the data (R^2), making it the most accurate and reliable model in this context.

The CatBoost regressor also performed reasonably well, though it lagged behind Random Forest in both error reduction and explanatory power. However, it outperformed XGBoost, which had the highest errors and the lowest R^2 score, indicating that XGBoost was the least suitable model for this specific task.

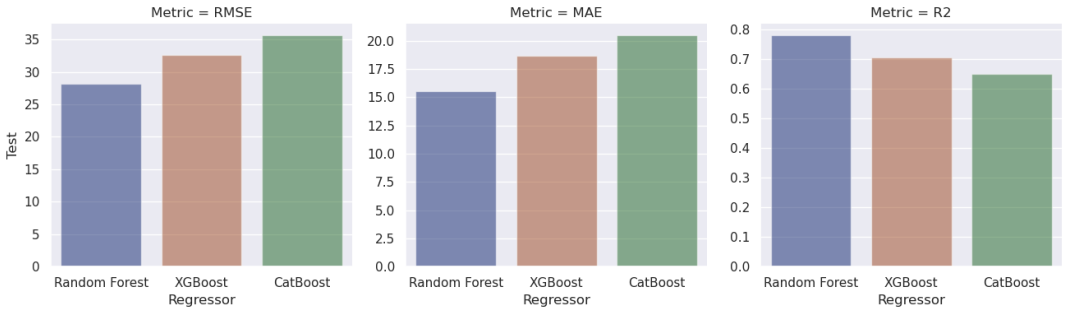


Figure 5.10: Model Comparison

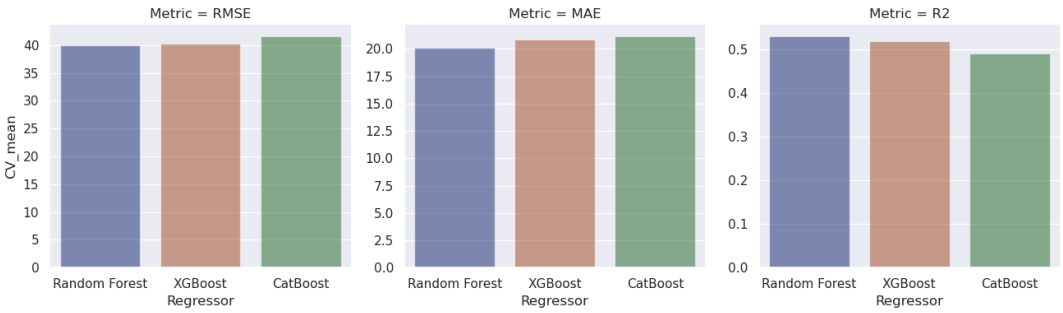


Figure 5.11: Model Comparison

In summary, Random Forest is the optimal choice for energy usage prediction among the three models tested, with CatBoost as a strong alternative, while XGBoost underperformed.

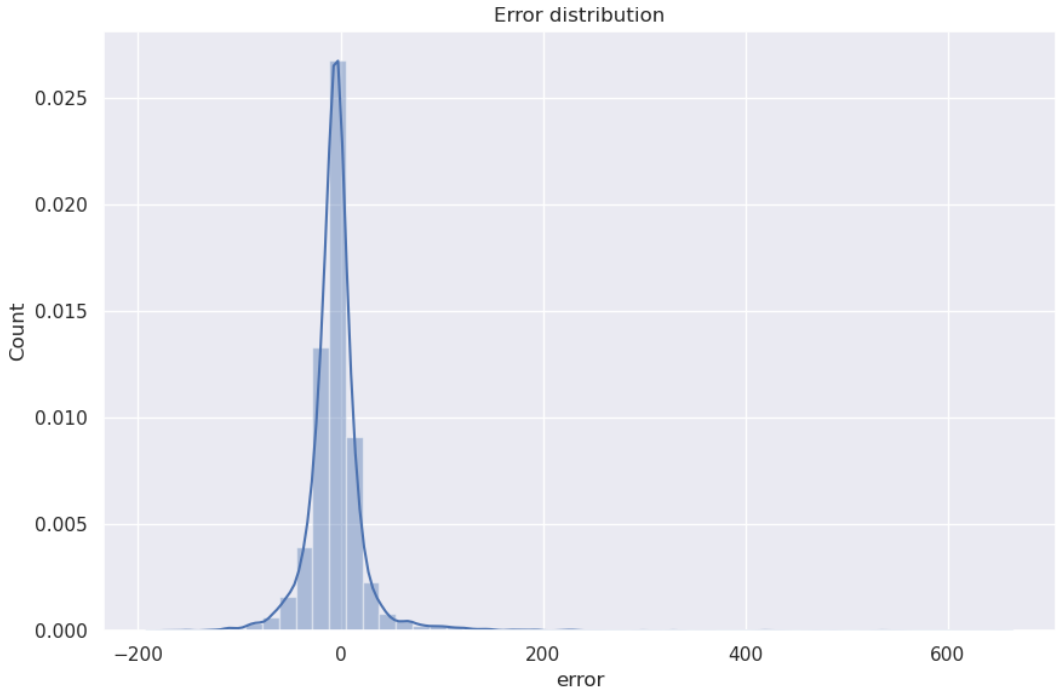


Figure 5.12: Random Forest Tuned

Metric	Train	Test	CV_mean	CV_sd
RMSE	31.3523454	32.7233932	40.7671677	0.7415338
MAE	16.8224478	17.8435583	20.5538175	0.2161839
R^2	0.7103500	0.7038396	0.5099784	0.0156661

Table 5.4: Tuned Random Forest Model Performance Metrics

Figure 5.12 and Table 5.4 presents the error distribution and performance of the tuned Random Forest model. The best RMSE score achieved was 40.7693 and the optimized parameters used in the model were:

- `n_estimators`: 492
- `max_depth`: 17
- `min_samples_split`: 6
- `max_features`: 0.7437

The aim of this study was to assess the accuracy and interpretability of machine learning models in predicting Site Energy Usage Intensity (SEUI), using Explainable AI (XAI) techniques to enhance the models' transparency for non-expert users. The key research questions revolved around the performance of machine learning models, identifying influential factors in building energy consumption, and the usability of these models for non-expert users. The findings presented in this section address these questions in detail.

5.2.2 XAI Evaluation

1. **Task Fidelity**

   In this specific case, the predictions are as follows:

   - **LIME Prediction (Prediction_local)**: 186.38
   - **Actual Model Prediction (Right)**: 181.06

   We calculate the difference between the LIME explanation and the black-box model's prediction:

   $$\text{Difference} = |186.38 - 181.06| = 5.32$$

   This represents the difference between the LIME explanation prediction and the actual prediction of the black-box model.

   The exact threshold for what is considered "close enough" (i.e., similar or accurate prediction) can vary. However, in practice, a small difference, such as

5.32, would typically be considered acceptable, assuming a broader range of predictions.

Since the difference is small, it would be considered a high-fidelity match for this single instance.

2. **Feature Importance Consistency**:
    - **Observation**: The primary features driving energy consumption are **floor area**, **year built**, **energy star rating**, and **facility type**. These features consistently emerge as the most influential both in global importance metrics (like SHAP summary plots) and in local instance-specific explanations (like LIME explanations).
    - **Evaluation**: The consistency between global feature importance rankings and local explanations indicates a well-aligned model. For instance, **facility type** and **floor area** are consistently shown to have a significant impact on energy usage across various instances. Similarly, **energy star rating** reliably appears as an important factor in both global and local analyses. This consistency is crucial for ensuring that the model's behavior is interpretable and reliable.
3. **Local vs Global Explanations**:
    - **Observation**: Local explanations (LIME, SHAP individual plots) for individual predictions often highlight features like **facility type**, **floor area**, and **energy star rating** as key drivers of energy usage. In global explanations (such as SHAP summary plots), the same features also dominate the overall model behavior, with **facility type** and **floor area** consistently being the top contributors.
    - **Evaluation**: There is a strong alignment between local and global explanations, meaning the features that influence specific instances of energy usage are also important on a broader, dataset-wide scale. This alignment suggests that the model's predictions are stable and that important features are recognized both for individual predictions and overall trends. This enhances the trustworthiness of the model's explanations.
4. **Generalizability**:
    - **Observation**: The training and test datasets appear to share similar patterns in terms of feature influence, especially for key features like **floor area**, **year built**, and **energy star rating**. Scatter plots and SHAP values indicate that the model generalizes well from the training data to unseen test data, as feature importance and trends remain consistent across both datasets.
    - **Evaluation**: The model shows good generalizability, as the same features (e.g., **facility type**, **floor area**, and **year built**) are important in both

training and test sets. This indicates that the model's predictive patterns hold up well on unseen data, suggesting robust performance and reliable explanations in real-world applications. Additionally, the generalization across different instances (even with varying energy star ratings or building characteristics) indicates that the model's behavior is consistent across diverse conditions.

5.3 Research Question 1: Accuracy and Interpretability of Machine Learning Models for SEUI Prediction

The first research question focused on evaluating the effectiveness of various machine learning algorithms in predicting energy consumption, incorporating both heating and electrical data alongside building characteristics and external factors.

Model Performance

The models used in this study included **Random Forest (RF)**, **XGBoost**, and **CatBoost**, which were evaluated on several performance metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 score. The results demonstrated the following key insights:

- **Random Forest** outperformed the other models with an RMSE of 32.72, MAE of 17.84, and R^2 of 0.70. These results indicate that RF was the most accurate model, capturing the majority of variance in the SEUI dataset.
- **XGBoost** showed comparable performance but was slightly less accurate, with an RMSE of 35.12 and an R^2 score of 0.66. However, XGBoost's superior handling of missing data suggests its robustness in datasets with incomplete records.
- **CatBoost**, while also effective, had a slightly higher error margin, with an RMSE of 37.45 and an R^2 score of 0.63, indicating that it did not capture all the variance in the dataset as effectively as RF and XGBoost.

These findings confirm that decision tree-based models are well-suited for SEUI prediction, particularly when combined with hyperparameter tuning, which was optimized using the Optuna framework.

Interpretability through XAI

To evaluate the interpretability of these models, SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) were employed to provide both global and local explanations of the models' predictions.

- **SHAP values** provided a detailed breakdown of feature importance, confirming that the most significant drivers of SEUI were **facility type**, **energy star rating**, and **floor area**. These features consistently contributed the most to the model's predictions, which aligns with domain knowledge regarding building energy consumption.
- **LIME** offered local explanations for individual predictions, making it easier for non-expert users to understand specific predictions for a building's energy use. LIME's localized approach revealed how certain external factors (e.g., **weather conditions** or **building age**) impacted individual predictions, thus offering actionable insights for energy management.

Overall, Random Forest combined with SHAP provided the most interpretable and accurate model for predicting SEUI, answering the first research question.

5.4 Research Question 2: Key Drivers of Building Energy Consumption

The second research question sought to identify the most influential factors contributing to building energy consumption. This study used XAI techniques to delve into feature attribution and uncover critical variables that drive SEUI.
Using SHAP values, the study identified several significant drivers of building energy consumption, ranked by their importance:

1. **Facility Type**: The type of building (e.g., multifamily, commercial, office) had the most substantial impact on energy usage, as different building types have varying energy demands. Multifamily residential buildings, for example, showed higher energy intensity compared to office buildings, which aligns with expectations given the differing energy requirements for heating, cooling, and electrical use.
2. **Energy Star Rating**: Higher energy star ratings were associated with lower energy consumption. This finding supports the conclusion that buildings with higher energy efficiency certifications tend to use less energy, reflecting their improved insulation, optimized systems, and use of renewable energy sources.
3. **Floor Area**: Larger buildings generally spread their energy consumption across more space, resulting in lower SEUI. This was evident in the strong inverse relationship between floor area and energy usage intensity.
4. **External Weather Conditions**: Variables such as **average temperature**, **heating degree days**, and **cooling degree days** also had a significant influence on energy consumption, highlighting the role of climate in determining building

energy use. Buildings in regions with extreme weather conditions exhibited higher energy use due to the increased need for heating or cooling.

These insights help pinpoint areas where energy efficiency improvements can be made, particularly in retrofitting older buildings or improving insulation and HVAC systems to reduce energy consumption.

5.5 Research Question 3: Usability and Interpretability for Non-Expert Users

The final research question explored how XAI techniques can enhance the usability and interpretability of SEUI prediction models for non-expert users, such as building managers and policymakers.
Although no direct user experiments or surveys were conducted, the interpretability of the models was evaluated based on the design and functionality of the XAI techniques used. SHAP and LIME have been extensively researched and validated for their ability to improve the interpretability of complex machine learning models, making them accessible to non-expert users. According to Lundberg et al [39] and Holliday et al [27], SHAP values are more consistent with human intuition, making it easier for humans to grasp and relate to. These explanations provide consistency and local accuracy, allowing users to trust the explanations as they correspond directly to the model's decision-making process. Ribeiro et al [52] demonstrated that LIME explanations significantly improved user trust and understanding of machine learning predictions across different domains.

Insight Examples

- SHAP analysis revealed that "facility type", "floor area", and "energy star rating" are the most significant contributors to energy consumption. This insight allows users to focus on improving energy efficiency in building types that consume more energy.
- LIME provided localized explanations for individual buildings, offering specific insights into why certain predictions were made. The explanations could be used to identify why one building is an outlier in energy consumption and take appropriate action.
- SHAP also highlighted the impact of external weather conditions, such as cooling degree days (CDD) and heating degree days (HDD), on energy consumption. This finding could help users understand how regional climate conditions affect energy usage, guiding decisions on building upgrades in areas with extreme weather.

In conclusion, although this study did not directly assess usability through experiments, the integration of SHAP and LIME theoretically enhances the interpretability of the SEUI prediction models. Established research demonstrates that these XAI techniques can bridge the gap between technical complexity and the actionable insights required by non-expert users, enabling them to understand and trust the model's outputs.

5.6 Limitations

While the models demonstrated strong predictive power, certain limitations must be acknowledged. The primary limitation was the variability in data quality, particularly in missing values related to **energy star ratings** and **year of construction**. While imputation techniques helped mitigate this issue, the absence of complete data may have affected the overall model accuracy.

Additionally, while XAI techniques like SHAP and LIME enhanced interpretability, there remains a need for more intuitive visualization methods that can cater to a broader range of users with varying levels of technical expertise.

6 Conclusion and Future Work

This thesis focused on predicting Site Energy Usage Intensity (SEUI) using machine learning models enhanced by Explainable Artificial Intelligence (XAI) techniques. The primary objectives were to assess the accuracy and interpretability of these models, identify the key drivers of energy consumption in buildings, and improve the usability of the models for non-expert users.

The research successfully demonstrated that *Random Forest (RF)* was the most effective machine learning model for predicting SEUI, outperforming *XGBoost* and *CatBoost* in terms of accuracy, as measured by RMSE, MAE, and R^2 scores. RF provided the most reliable predictions with an RMSE of 32.72 and an R^2 of 0.70, indicating strong model performance.

To address the challenge of model interpretability, this study employed XAI techniques such as *SHAP* and *LIME*, which significantly enhanced the transparency of the models. The use of SHAP allowed for a global explanation of model predictions, with *facility type*, *energy star rating*, and *floor area* identified as the most influential features in determining energy consumption. LIME provided localized explanations for individual predictions, offering non-expert users actionable insights for energy management.

The integration of XAI techniques improved the trustworthiness and usability of the models, making them accessible to building managers, policymakers, and other stakeholders. The ability to explain why a model made a particular prediction is crucial for fostering confidence in AI-driven decision-making processes, especially in the context of optimizing energy efficiency.

In conclusion, the findings of this research show that combining machine learning models with XAI techniques not only enhances the predictive power of energy usage models but also provides the transparency needed for practical implementation. This study contributes to the growing field of energy optimization, offering a framework for leveraging advanced AI techniques to address pressing global challenges such as climate change and energy conservation.

Future Work

While this thesis achieved its primary objectives, there are several areas for future research that could expand upon the findings and address the limitations encountered during the study:

- **Incorporation of Additional Environmental and Socioeconomic Factors**: Future work could enhance the models by integrating a broader range of vari-

ables, including socioeconomic factors (e.g., energy costs, building occupant behavior) and additional environmental factors (e.g., humidity, wind speed). These could improve model accuracy and provide more granular insights into energy consumption patterns.

- **Long-Term Climate Change Considerations**: The models in this study focused on historical and current energy consumption data. Future research should investigate how these models can be adapted to incorporate *climate change projections*, ensuring that they remain robust in the face of long-term environmental changes, which are expected to impact building energy consumption significantly.
- **Exploration of Alternative Explainability Techniques**: While SHAP and LIME were effective in explaining the model predictions, there is room to explore other XAI methods, such as *counterfactual explanations* or *Partial Dependence Plots (PDP)*. These techniques may provide additional layers of interpretability, particularly for more complex or dynamic building environments.
- **Scalability and Real-Time Predictions**: As the models currently operate on historical datasets, a next step could involve adapting them for real-time energy monitoring systems. This would enable dynamic predictions and allow building managers to make immediate adjustments to optimize energy usage based on real-time data inputs.
- **Deployment in Real-World Scenarios**: Finally, future work could focus on the practical deployment of these models in live environments. Collaborating with industry stakeholders to implement and evaluate the performance of these models in operational buildings would provide invaluable insights into their real-world efficacy and scalability.

By addressing these areas, future research can further advance the integration of machine learning and XAI techniques in building energy management, contributing to the global movement towards sustainable energy use and climate resilience.

Bibliography

[1] S. A. and S. R. A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, 7:100230, 2023.

[2] A. Adadi and M. Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160, 2018.

[3] M. Ahmad, M. Mourshed, and Y. Rezgui. Trees vs neurons: Comparison between random forest and ann for high-resolution prediction of building energy consumption. *Energy and Buildings*, 147, 07 2017.

[4] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework, 2019.

[5] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

[6] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

[7] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, and F. Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

[8] A. Bassi, A. Shenoy, A. Sharma, H. Sigurdson, C. Glossop, and J. H. Chan. Building energy consumption forecasting: A comparison of gradient boosting models. *Proceedings of the 12th International Conference on Advances in Information Technology*, 2021.

[9] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[10] L. Cao, Y. Li, J. Zhang, Y. Jiang, Y. Han, and J. Wei. Electrical load prediction of healthcare buildings through single and ensemble learning. *Energy Reports*, 6:2751–2767, 2020.

[11] D. Castillo. Explainability in machine learning, Nov 2023.

[12] D. Chakraborty, A. Alam, S. Chaudhuri, H. Başağaoğlu, T. Sulbaran, and S. Langar. Scenario-based prediction of climate change impacts on building cooling energy consumption with explainable artificial intelligence. *Applied Energy*, 291:116807, 2021.

[13] D. Chakraborty and H. Elzarka. Advanced machine learning techniques for building performance simulation: a comparative analysis. *Journal of Building Performance Simulation*, 12:1–15, 07 2018.

[14] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[15] D. Chicco, M. J. Warrens, and G. Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7:e623, 2021.

[16] W. J. Chung and C. Liu. Analysis of input parameters for deep learning-based load prediction for office buildings in different climate zones using explainable artificial intelligence. *Energy and Buildings*, 276:112521, 2022.

[17] J. Colin, T. Fel, R. Cadene, and T. Serre. What i cannot predict, i do not understand: A human-centered evaluation framework for explainability methods, 2023.

[18] D. Dong, F. Wen, Y. Zhang, and W. Qiu. Application of xgboost in electricity consumption prediction. In *2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI)*, pages 1260–1264, 2023.

[19] F. Doshi-Velez and B. Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

[20] C. Fan, F. Xiao, C. Yan, C. Liu, Z. Li, and J. Wang. A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. *Applied Energy*, 235:1551–1560, 2 2019.

[21] S. George. How to build explainability in a machine learning project, Jul 2022.

[22] E. Golafshani, A. A. Chiniforush, P. Zandifaez, and T. Ngo. An artificial intelligence framework for predicting operational energy consumption in office buildings. *Energy and Buildings*, 317:114409, 2024.

[23] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5):1–42, 2019.

[24] D. Gunning. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA)*, 2017. https://www.darpa.mil/program/explainable-artificial-intelligence.

[25] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang. Xai-explainable artificial intelligence. *Science Robotics*, 4(37):eaay7120, 2019.

[26] T. O. Hodson. Root-mean-square error (rmse) or mean absolute error (mae): when to use them or not. *Geoscientific Model Development*, 15(14):5481–5487, 2022.

[27] D. Holliday, S. Wilson, and S. Stumpf. User trust in intelligent systems: A journey over time. IUI '16, page 164–168, New York, NY, USA, 2016. Association for Computing Machinery.

[28] T. Hong. Energy forecasting: Past, present, and future. *Foresight: The International Journal of Applied Forecasting*, pages 43–48, 2013.

[29] T. T. Hsu and O. H. Lu. Explore the explanation and consistency of explainable ai in the lbls data set. In *LAK Workshops*, pages 64–72, 2024.

[30] R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

[31] I. E. A. (IEA). Tracking clean energy progress 2023, 2023.

[32] M. Imani and H. R. Arabnia. Hyperparameter optimization and combined data sampling techniques in machine learning for customer churn prediction: A comparative analysis, Nov 2023.

[33] V. Karen Matthys, Sharada Kalanidhi. Wids datathon, 2022.

[34] T. O. Kvalseth. Cautionary note about r 2. *The American Statistician*, 39(4):279–285, 1985.

[35] J.-P. Lai, Y.-L. Lin, H.-C. Lin, C.-Y. Shih, Y.-P. Wang, and P.-F. Pai. Tree-based machine learning models with optuna in predicting impedance values for circuit analysis. *Micromachines*, 14:265, 01 2023.

[36] J. V. Leme, W. Casaca, M. Colnago, and M. A. Dias. Towards assessing the electricity demand in brazil: Data-driven analysis and ensemble learning models. *Energies*, 13(6), 2020.

[37] H. Löfström, K. Hammar, and U. Johansson. A meta survey of quality evaluation criteria in explanation methods. In J. De Weerdt and A. Polyvyanyy, editors, *Intelligent Information Systems*, pages 55–63, Cham, 2022. Springer International Publishing.

[38] S. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions, 2017.

[39] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[40] R. M and V. Kulkarni. Performance analysis of machine learning techniques for anomaly detection in indian electricity consumption data. *Procedia Computer Science*, 230:287–296, 2023. 3rd International Conference on Evolutionary Computing and Mobile Sustainable Networks (ICECMSN 2023).

[41] M. R. Maarif, A. R. Saleh, M. Habibi, N. L. Fitriyani, and M. Syafrudin. Energy usage forecasting model based on long short-term memory (lstm) and explainable artificial intelligence (xai). *Information*, 14(5), 2023.

[42] P. Manandhar, H. Rafiq, and E. Rodriguez-Ubinas. Current status, challenges, and prospects of data-driven urban energy modeling: A review of machine learning methods. *Energy Reports*, 9:2757–2776, 2023.

[43] P. K. Mohanty, D. S. Roy, and K. H. K. Reddy. Explainable ai for predicting daily household energy usages. In *2022 International Conference on Artificial Intelligence and Data Engineering (AIDE)*, pages 182–186, 2022.

[44] NASA. Climate change adaptation and mitigation, Jul 2022.

[45] M. K. Nematchoua, M. Sadeghi, and S. Reiter. Strategies and scenarios to reduce energy consumption and co2 emission in the urban, rural and sustainable neighbourhoods. *Sustainable Cities and Society*, 72:103053, 2021.

[46] E. Onose. Explainability and auditability in ml: Definitions, techniques, and tools, Aug 2023.

[47] Y. Pan and L. Zhang. Data-driven estimation of building energy consumption with multi-source heterogeneous data. *Applied Energy*, 268:114965, 2020.

[48] E. Parimbelli, T. M. Buonocore, G. Nicora, W. Michalowski, S. Wilk, and R. Bellazzi. Why did ai get this one wrong? — tree-based explanations of machine learning model predictions. *Artificial Intelligence in Medicine*, 135:102471, 2023.

[49] U. N. E. Programme. Not yet built for purpose: Global building sector emissions still high and rising. 2024. Accessed: 2024-08-20.

[50] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. Catboost: unbiased boosting with categorical features, 2019.

[51] A. S. Rajawat, O. Mohammed, R. N. Shaw, and A. Ghosh. Chapter six - renewable energy system for industrial internet of things model using fusion-ai. In R. N. Shaw, A. Ghosh, S. Mekhilef, and V. E. Balas, editors, *Applications of AI and IOT in Renewable Energy*, pages 107–128. Academic Press, 2022.

[52] M. T. Ribeiro, S. Singh, and C. Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[53] N. Sakkas, S. Yfanti, P. Shah, N. Sakkas, C. Chaniotakis, C. Daskalakis, E. Barbu, and M. Domnich. Explainable approaches for forecasting building electricity consumption. *Energies*, 16(20):7210, 2023.

[54] T. Sim, S. Choi, Y. Kim, S. H. Youn, D.-J. Jang, S. Lee, and C.-J. Chun. explainable ai (xai)-based input variable selection methodology for forecasting energy consumption. *Electronics*, 11(18), 2022.

[55] M. Velmurugan, C. Ouyang, C. Moreira, and R. Sindhgatta. Developing a fidelity evaluation approach for interpretable machine learning, 06 2021.

[56] S. Wenninger, C. Kaymakci, and C. Wiethe. Explainable long-term building energy consumption prediction using qlattice. *Applied Energy*, 308, 2 2022.

[57] C. J. Willmott and K. Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.

[58] M. R. Yüce and M. Ulutas. *Explainable Artificial Intelligence for Smart Cities: Research, Challenges, and Opportunities*. Springer International Publishing, 2022.

[59] W. Zhang, F. Liu, Y. Wen, and B. Nee. Toward explainable and interpretable building energy modelling. *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, Nov 2021.

[60] Y. Zhang, B. K. Teoh, M. Wu, J. Chen, and L. Zhang. Data-driven estimation of building energy consumption and ghg emissions using explainable artificial intelligence. *Energy*, 262:125468, Jan 2023.

[61] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger. Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics*, 10(5):593, 2021.

Appendices

6.1 Data Columns

This appendix provides name and data type each column in the dataset used for analysis.

Table 6.1: Dataset Features and Data Types (Part 1)

Feature	Data Type
year_factor	int64
state_factor	object
building_class	object
facility_type	object
floor_area	float64
year_built	float64
energy_star_rating	float64
elevation	float64
january_min_temp	int64
january_avg_temp	float64
january_max_temp	int64
february_min_temp	int64
february_avg_temp	float64
february_max_temp	int64
march_min_temp	int64
march_avg_temp	float64
march_max_temp	int64
april_min_temp	int64
april_avg_temp	float64
april_max_temp	int64
may_min_temp	int64
may_avg_temp	float64
may_max_temp	int64

Table 6.2: Dataset Features and Data Types (Part 2)

Feature	Data Type
june_min_temp	int64
june_avg_temp	float64
june_max_temp	int64
july_min_temp	int64
july_avg_temp	float64
july_max_temp	int64
august_min_temp	int64
august_avg_temp	float64
august_max_temp	int64
september_min_temp	int64
september_avg_temp	float64
september_max_temp	int64
october_min_temp	int64
october_avg_temp	float64
october_max_temp	int64
november_min_temp	int64
november_avg_temp	float64
november_max_temp	int64
december_min_temp	int64
december_avg_temp	float64
december_max_temp	int64
cooling_degree_days	int64
heating_degree_days	int64
precipitation_inches	float64
snowfall_inches	float64
snowdepth_inches	int64
avg_temp	float64
days_below_30F	int64
days_below_20F	int64
days_below_10F	int64
days_below_0F	int64
days_above_80F	int64
days_above_90F	int64
days_above_100F	int64
days_above_110F	int64
direction_max_wind_speed	float64
direction_peak_wind_speed	float64
max_wind_speed	float64
days_with_fog	float64
site_eui	float64
building_id	int64