Course Project

UNDERSTANDING AND FORECASTING ELECTRICITY DEMAND

Name: Lidia Y. Bereketeab - 20029

Name : Arsiema Y. Gebremichael -20039

Section: CS481

Instructor: Dr. Beketayev, Kenes

Date: March 27th, 2024

INTRODUCTION 3

Database description 3

Tasks 4

ANALYSIS 5

Identifying External Influences on Electricity Demand 5
Correlation Analysis 5
Basic Regression Analysis 6
Tocumen 6
Santiago 7
David 8
Visualization Techniques 10
Heatmap 10
Scatterplot 11
Histogram 12
Boxplot 13

Managing Uncertainty in Electricity Load Forecasting 14
Uncertainty Analysis (Bootstrapping): 14
Scenario Analysis (Optimistic and Pessimistic Scenarios): 16
Sensitivity Analysis: 17
Shapely method for building the sensitivity analysis. 19
Model-specific Shapley Values: 19
Temporal Shapley Values: 24
Simulation 26
Monte Carlo simulations 26
Discrete-event simulations 28
Discontinuity regression 30
Regression Discontinuity Design (RDD): 30
Segmented Regression Analysis: 34
Explainability of variables 36
SHAP Values (SHapley Additive exPlanations): 37
LIME (Local Interpretable Model-agnostic Explanations): 38
Feature importance and Decision Tree. 40
Conclusion 41

Understanding Seasonal Patterns in Electricity Consumption 42
FB PROPHET: 42
Introduction 42
Advantages of Facebook Prophet 42
Analysis Overview 42
Autoregressive Model Analysis for Electricity Demand Forecasting 48
How Autoregressive Models Work 48
Autocorrelation Analysis 48

Conclusion: 51

CONCLUSION **52**

REFERENCES **54**

Introduction

Electricity is the lifeblood of modern society, powering everything from our homes to our industries. The ability to accurately predict electricity demand is crucial for ensuring a reliable and efficient energy supply. Our project, "Understanding and Forecasting Electricity Demand," delves into the complexities of this challenge, aiming to unravel the seasonal patterns, external influences, and inherent uncertainties surrounding electricity consumption.

Electricity load forecasting is a critical task for energy planners, policymakers, and utility providers. It involves predicting the future demand for electricity based on historical data, current trends, and various influencing factors. Accurate forecasts enable stakeholders to make informed decisions regarding infrastructure investments, energy generation, and resource allocation.

In this project, we focus on analyzing historical electricity load data to identify seasonal patterns, understand the impact of external factors such as weather conditions, and assess the sources of uncertainty in electricity load forecasting. By leveraging techniques ranging from statistical analysis to machine learning, we aim to develop robust forecasting models that can support energy planning and infrastructure decisions.

Database description

- nat_demand: National electricity load
- T2M: Temperature at 2 meters
- QV2M: Relative humidity at 2 meters
- TQL: Liquid precipitation
- W2M: Wind speed at 2 meters

And after the underscore is the city

- toc: Tocumen, Panama city
- san: Santiago city
- dav: David city

The rest of variables:

- Holiday_ID: Unique identification number integer
- holiday: Holiday binary indicator (1=holiday, 0=regular day)
- school: School period binary indicator (1=school, 0=vacations)

Tasks

Through our investigation, we seek to address fundamental questions such as the following:

1. What external factors influence electricity demand, and to what extent?
2. What are the sources of uncertainty in electricity load forecasting, and how can we mitigate them to improve the accuracy of our predictions?
3. What are the seasonal patterns in electricity consumption, and how do they vary over time?

By exploring these questions, we endeavor to contribute to the development of more reliable and efficient methods for forecasting electricity demand, ultimately aiding in the sustainable management of energy resources. Join us as we embark on this journey to unlock insights into one of the most critical aspects of modern civilization – the demand for electricity.

Analysis

Identifying External Influences on Electricity Demand

> Description: Investigate the impact of external factors, such as weather conditions, on electricity demand. Explore whether changes in weather (e.g., temperature fluctuations) coincide with changes in electricity consumption. Use basic methods to check the correlation between these external factors and electricity use.

Correlation Analysis

The correlation coefficients indicate the strength and direction of the linear relationship between 'nat_demand' and each factor. A correlation coefficient close to 1 or -1 suggests a strong positive or negative linear relationship respectively, while a coefficient close to 0 indicates a weak linear relationship.

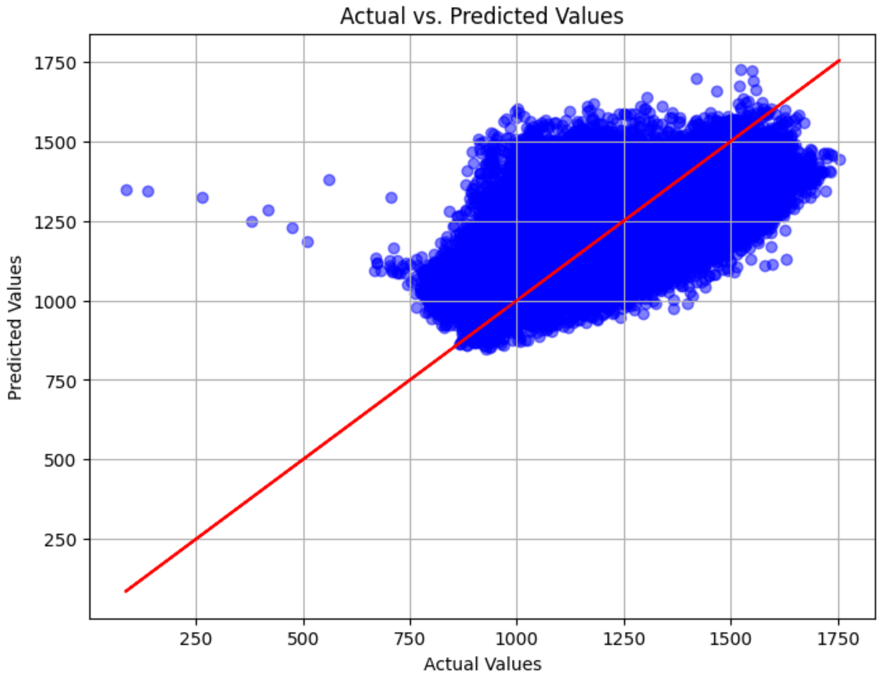
Factors with Strong Correlation: Factors with higher absolute correlation coefficients are more strongly associated with 'nat_demand'. In this analysis, 'T2M_toc', 'T2M_san', and 'T2M_dav' (presumably representing temperature in different locations) exhibit relatively high positive correlations with 'nat_demand'. This suggests that temperature is a significant driver of electricity demand, with higher temperatures correlating with increased demand.

Factors with Weak Correlation: Factors with correlation coefficients closer to 0 have weaker associations with 'nat_demand'. For instance, 'QV2M_toc', 'QV2M_san', 'QV2M_dav', 'TQL_toc', and 'TQL_dav' have correlations close to 0, indicating that specific humidity and total cloud liquid water content have minimal linear relationship with national electricity demand.

Holiday and School Variables: The 'Holiday_ID', 'holiday', and 'school' variables show negative correlations with 'nat_demand', albeit relatively weak. This suggests that holidays and school schedules may have a slight dampening effect on electricity demand, possibly due to reduced commercial and industrial activities during holidays and non-school days.

Basic Regression Analysis

Tocumen



T2M_toc, QV2M_toc, TQL_toc, W2M_toc:

R-squared: The R-squared value is 0.431, indicating that approximately 43.1% of the variance in national electricity demand (nat_demand) can be explained by the variation in the independent variables (T2M_toc, QV2M_toc, TQL_toc, W2M_toc).

Coefficients:

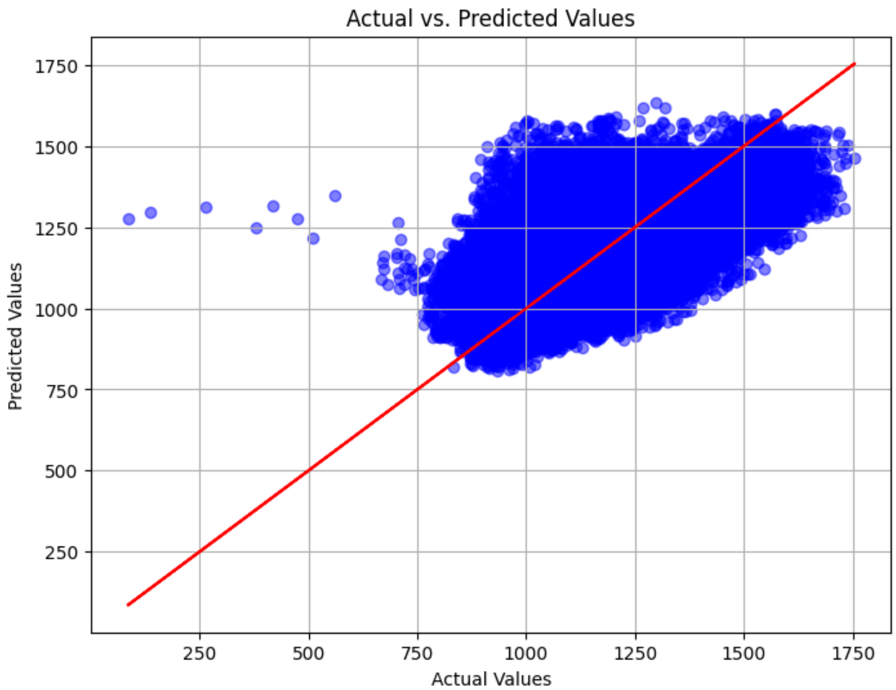
T2M_toc: The coefficient is 73.9871, suggesting that a one-unit increase in temperature at the specified location leads to an increase of approximately 74 units in national electricity demand, holding other variables constant.

QV2M_toc: The coefficient is -17.3490, but the p-value is not significant (0.973), indicating that specific humidity (QV2M_toc) does not have a significant linear relationship with national electricity demand.

TQL_toc: The coefficient is 92.1655, indicating that an increase in total cloud liquid water content leads to an increase in national electricity demand. The p-value is significant ($p < 0.05$), suggesting a significant linear relationship.

W2M_toc: The coefficient is 2.0165, indicating that an increase in wind speed leads to a slight increase in national electricity demand. The p-value is significant ($p < 0.05$), suggesting a significant linear relationship.

Santiago



T2M_san, QV2M_san, TQL_san, W2M_san:

R-squared: The R-squared value is 0.436, indicating that approximately 43.6% of the variance in national electricity demand can be explained by the variation in the independent variables at a different location.

Coefficients:

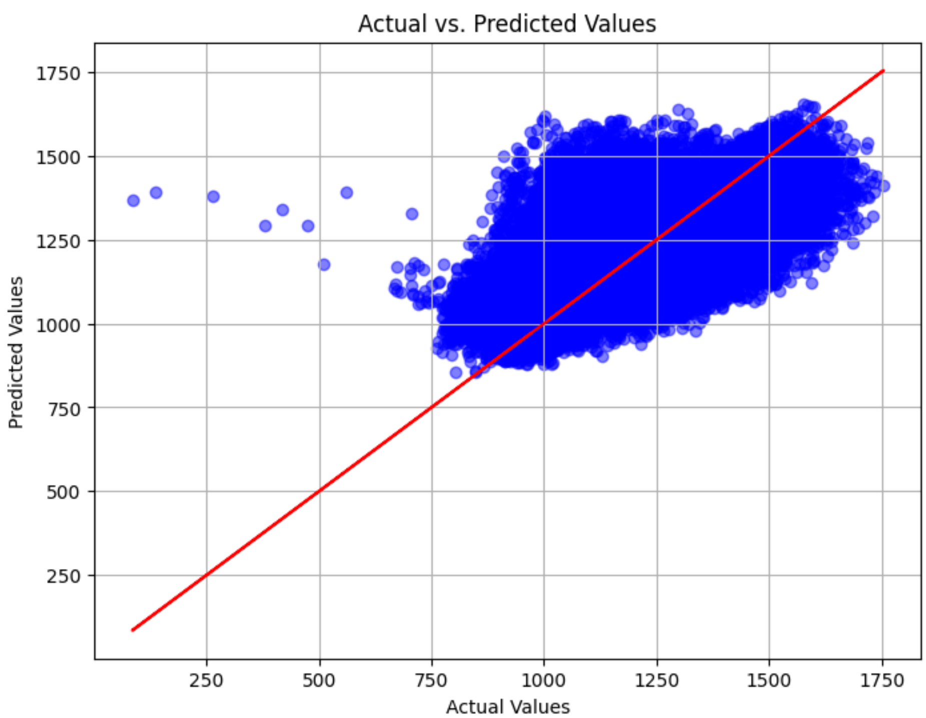
T2M_san: The coefficient is 41.0434, suggesting that temperature at this location positively influences national electricity demand.

QV2M_san: The coefficient is 23330, indicating a significant positive relationship between specific humidity and national electricity demand.

TQL_san: The coefficient is 9.6976, but the p-value is not significant (0.287), suggesting that total cloud liquid water content (TQL_san) does not have a significant linear relationship with national electricity demand.

W2M_san: The coefficient is 6.2440, indicating a significant positive relationship between wind speed and national electricity demand.

David



T2M_dav, QV2M_dav, TQL_dav, W2M_dav:

R-squared: The R-squared value is 0.442, indicating that approximately 44.2% of the variance in national electricity demand can be explained by the variation in the independent variables at another location.

Coefficients:

T2M_dav: The coefficient is 54.7151, suggesting a positive relationship between temperature at this location and national electricity demand.

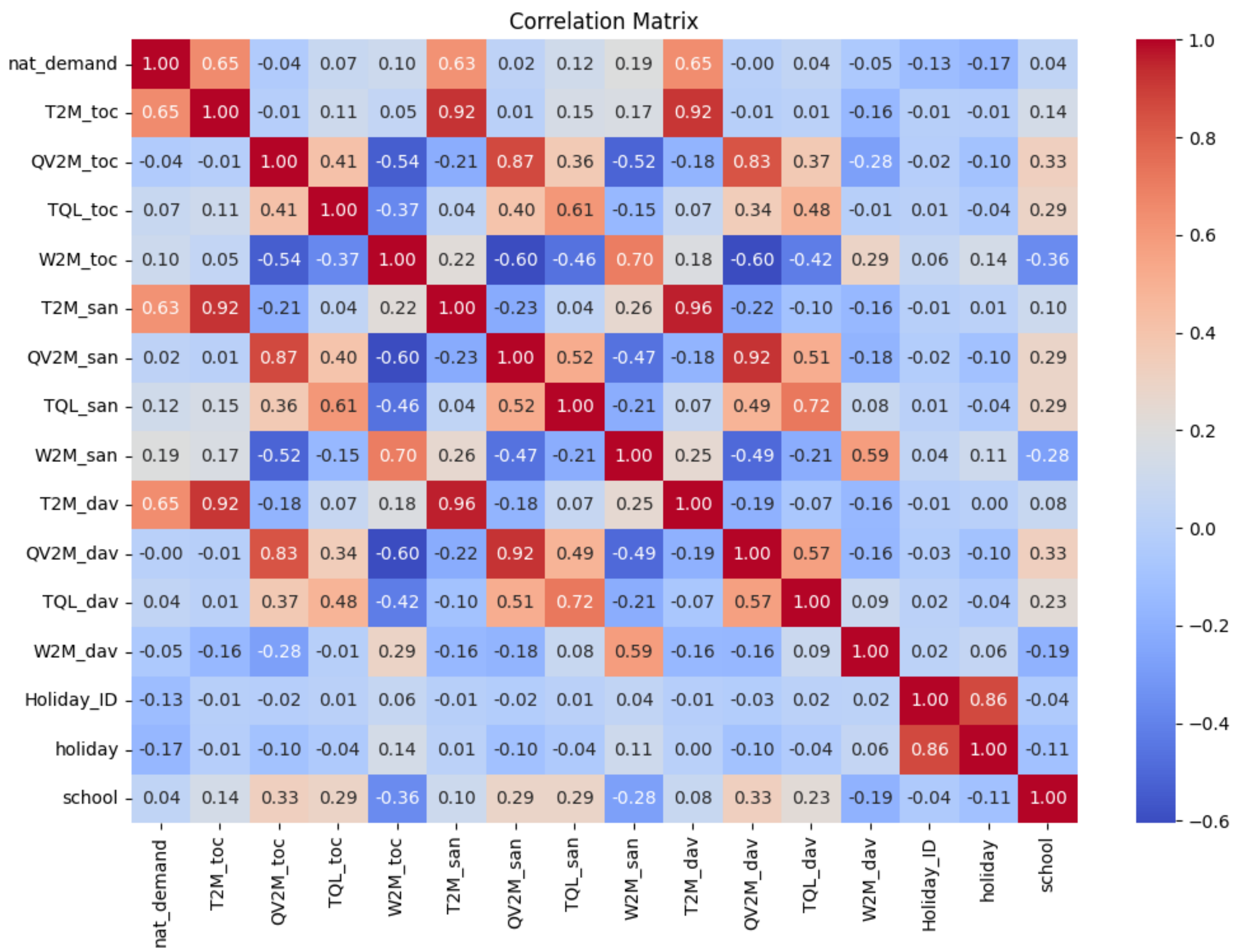
QV2M_dav: The coefficient is 17200, indicating a significant positive relationship between specific humidity and national electricity demand.

TQL_dav: The coefficient is 7.4946, but the p-value is not significant (0.423), suggesting that total cloud liquid water content (TQL_dav) does not have a significant linear relationship with national electricity demand.

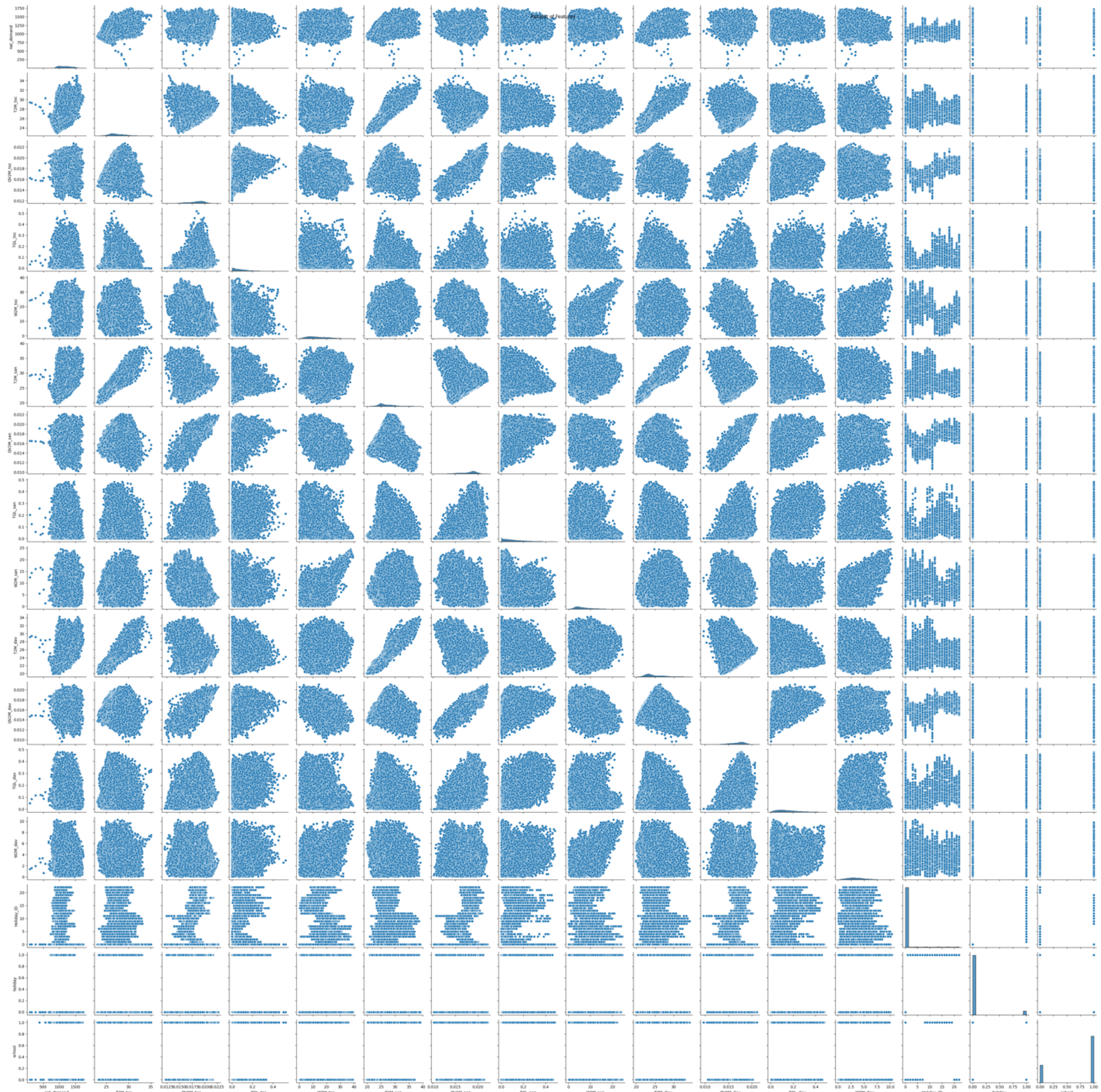
W2M_dav: The coefficient is 8.3536, indicating a significant positive relationship between wind speed and national electricity demand.

Visualization Techniques

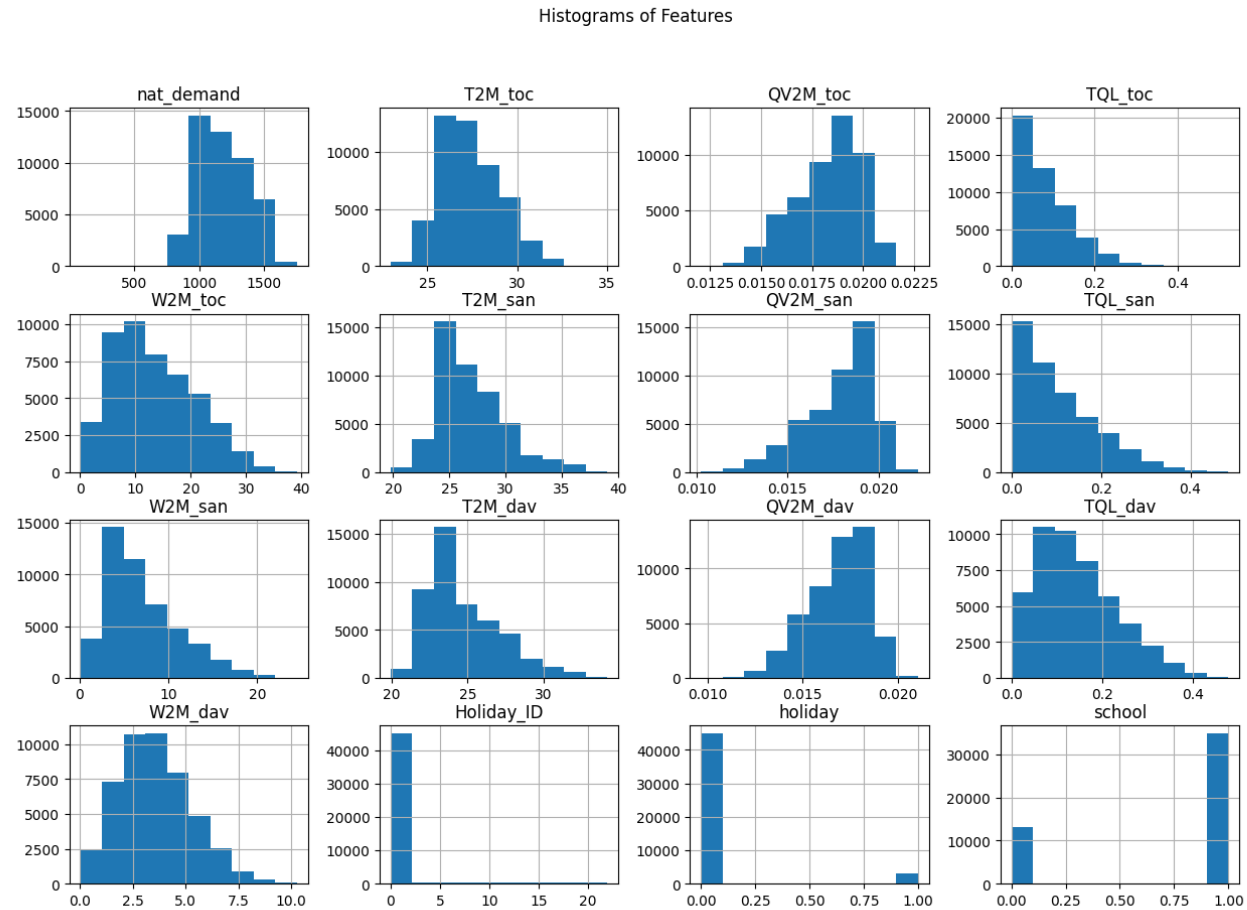
Heatmap



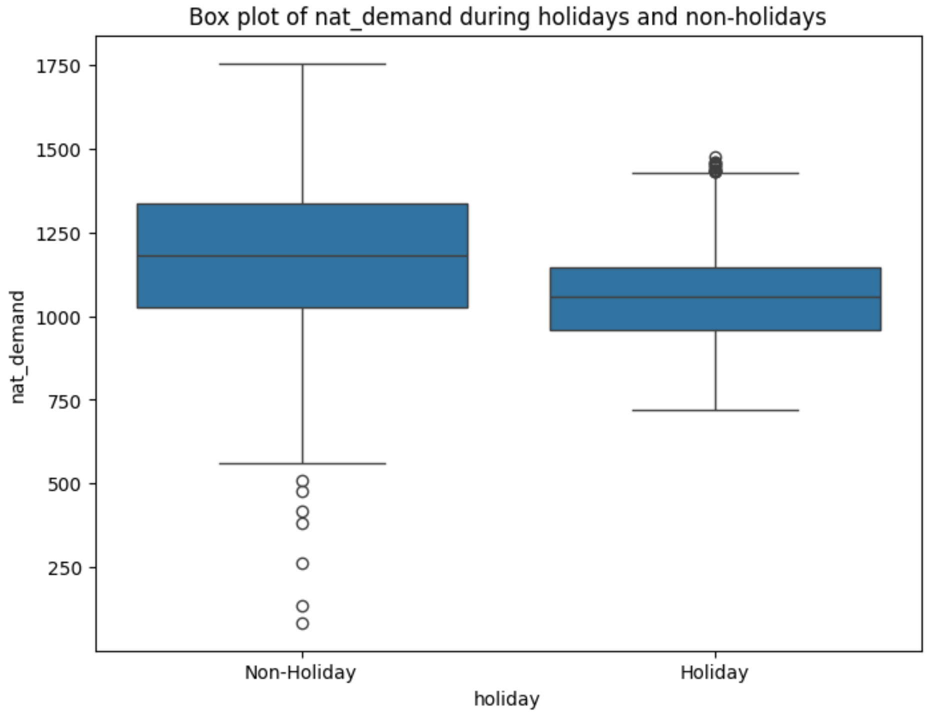
Scatterplot



Histogram


Histograms of Features
nat_demand
T2M_toc
QV2M_toc
TQL_toc
W2M_toc
T2M_san
QV2M_san
TQL_san
W2M_san
T2M_dav
QV2M_dav
TQL_dav
W2M_dav
Holiday_ID
holiday
school

Boxplot



Explanation:

The correlation matrix reveals the relationships between different variables. Among them, there are notable positive correlations, such as between 'nat_demand' and 'T2M_toc' (0.65), 'T2M_toc' and 'T2M_san' (0.92), and 'T2M_san' and 'T2M_dav' (0.96). These correlations suggest strong linear associations between these pairs of variables. Conversely, there are negative correlations present, like between 'QV2M_toc' and 'QV2M_san' (-0.54) and 'W2M_toc' and 'W2M_san' (0.70), indicating inverse relationships. Notably, some variables show weak or negligible correlations, such as 'nat_demand' with 'QV2M_toc' (-0.04), implying little linear relationship between natural demand and specific humidity over Tokyo. Additionally, the presence of strong correlations between 'holiday' and 'Holiday_ID' (0.86) could indicate redundancy or multicollinearity in the data, potentially influencing predictive models or analyses involving these variables. Overall, this correlation matrix provides valuable insights into the interdependencies within the dataset, guiding further analysis or modeling decisions.

Managing Uncertainty in Electricity Load Forecasting

> Description: Assess the potential risks and uncertainties associated with forecasting future electricity demand. Consider variability in input data, model assumptions, and external factors that could affect the accuracy of forecasts. Develop strategies to mitigate these risks and uncertainties to make more reliable predictions.

Uncertainty Analysis (Bootstrapping):

Bootstrapping is a technique used to assess this uncertainty by generating multiple samples from the original dataset. Each bootstrapped sample represents a possible realization of the historical data, capturing the inherent variability.

By calculating summary statistics (mean, standard deviation, confidence interval) from the forecast samples derived from bootstrapped datasets, we gain insights into the range of possible outcomes and the level of uncertainty associated with the forecasts. This helps in understanding the variability in electricity demand predictions and the potential risks involved.

In our output:

The **forecast mean (1182.85 in this case)** represents the average forecasted electricity demand across all bootstrapped samples. It serves as the central estimate or expected value of the forecast.

While the forecast mean provides a point estimate of the expected demand, it may not capture the full range of possible outcomes, especially in scenarios with significant variability or uncertainty.

The **forecast standard deviation (0.88 in this case)** quantifies the dispersion or spread of the forecasted demand values around the mean.

A higher standard deviation indicates greater variability or uncertainty in the forecasts, suggesting that the actual demand could deviate more substantially from the mean estimate.

In the context of electricity load forecasting, a higher standard deviation implies increased uncertainty in predicting future demand, which may pose challenges for decision-making and resource planning.

The 95% confidence interval ([1181.21, 1184.57] in this case) provides a range within which we are 95% confident the true forecast lies.

This interval accounts for the variability and uncertainty inherent in the forecasting process, capturing the potential range of outcomes with a specified level of confidence.

A wider confidence interval indicates higher uncertainty in the forecast, as it suggests a broader range of possible demand values.

Stakeholders can use the confidence interval to assess the reliability and precision of the forecast and make decisions considering the range of possible outcomes.

Risk Mitigation Strategies:

Data Quality Improvement: Use uncertainty analysis findings to prioritize data quality enhancement efforts. Invest in data validation, cleansing, and augmentation techniques to reduce uncertainties stemming from poor data quality.

Model Robustness Enhancement: Incorporate uncertainty quantification techniques into forecasting models to produce more robust predictions. Develop ensemble models that consider a range of potential scenarios to mitigate the impact of forecast uncertainties.

Dynamic Resource Allocation: Leverage uncertainty analysis results to inform resource allocation strategies. Implement dynamic resource allocation mechanisms that can adapt to changing forecast uncertainties, ensuring optimal resource utilization under uncertain conditions.

Scenario Analysis (Optimistic and Pessimistic Scenarios):

Uncertainty in electricity load forecasting also stems from future scenarios that may deviate from expected conditions. By defining optimistic and pessimistic scenarios and performing forecasting under these scenarios, we explore the range of possible outcomes under different conditions.

In our output:

The optimistic forecast (1301.16 in this case) represents the forecasted electricity demand under conditions where factors influencing demand are assumed to be more favorable or beneficial than expected.

This forecast reflects a scenario where demand may exceed expectations due to factors such as favorable economic conditions, increased consumer confidence, or unanticipated events driving higher consumption.

The optimistic forecast serves as an upper bound or best-case scenario, providing insight into the potential upside or opportunities associated with electricity demand projections.

The pessimistic forecast (1064.58 in this case) represents the forecasted electricity demand under conditions where factors influencing demand are assumed to be less favorable or detrimental than expected.

This forecast reflects a scenario where demand may fall short of expectations due to factors such as adverse economic conditions, regulatory changes, or unexpected disruptions leading to reduced consumption.

The pessimistic forecast serves as a lower bound or worst-case scenario, highlighting the potential downside or risks associated with electricity demand projections.

In summary, the output of the scenario analysis using optimistic and pessimistic scenarios provides insights into the range of possible outcomes for electricity load forecasting under different conditions. By considering both optimistic and pessimistic forecasts, stakeholders can assess the potential variability and uncertainty in demand projections, helping inform decision-making and risk management strategies. These scenarios allow stakeholders to explore the potential implications of different future conditions and prepare accordingly, thereby enhancing the resilience and adaptability of energy planning and management efforts.

Risk Mitigation Strategies:

Contingency Planning: Use scenario analysis insights to develop contingency plans tailored to different forecast scenarios. Establish protocols for responding to extreme events or deviations from expected conditions, ensuring resilience in the face of uncertainty.

Scenario-Based Decision-Making: Incorporate scenario analysis results into decision-making processes. Adopt scenario-based decision-making frameworks that account for the range of possible outcomes and prioritize actions based on their expected impact under different scenarios.

Stress Testing and Resilience Building: Use scenario analysis as a tool for stress testing and resilience building. Identify vulnerabilities and weaknesses in existing systems through scenario simulations and implement measures to enhance system resilience and adaptability.

Sensitivity Analysis:

Sensitivity analysis assesses the impact of changes in input variables on the forecasted outcomes. In electricity load forecasting, input variables such as weather conditions, economic factors, and policy changes can significantly influence demand predictions.

By systematically varying each input variable and observing the resulting changes in forecasted electricity demand, we identify the key drivers of uncertainty and assess their relative importance.

Understanding the sensitivity of the forecast to changes in input variables helps in identifying critical factors that may contribute to forecast uncertainty and informs strategies to mitigate risks.

In our output:

The sensitivity analysis evaluates the impact of each individual variable on the forecasted electricity load. It systematically varies each numerical input variable (excluding datetime columns) by increasing it by 10% and computes the corresponding forecasted load. The output provides the forecasted demand under each perturbed variable.

For example, for the 'nat_demand' column, where no perturbation is applied, the forecasted demand remains the same as the mean forecast. This suggests that changes in natural demand factors have minimal impact on the overall forecast.

Similarly, for other variables such as temperature ('T2M_toc', 'T2M_san', 'T2M_dav'), humidity ('QV2M_toc', 'QV2M_san', 'QV2M_dav'), and cloud cover ('TQL_toc', 'TQL_san', 'TQL_dav'), the forecasted demand also remains close to the mean forecast. These results indicate a relatively low sensitivity to changes in these variables.

Risk Mitigation Strategies:

Variable-Specific Risk Mitigation: Identify variables with the highest sensitivity to forecast uncertainties and develop targeted risk mitigation strategies for these variables. Implement data validation procedures, model calibration techniques, or external data integration strategies to reduce risks associated with sensitive variables.

Model Refinement and Validation: Use sensitivity analysis results to refine and validate forecasting models. Incorporate variable interactions and nonlinear relationships revealed by sensitivity analysis into model structures, improving model accuracy and reducing uncertainty.

Continuous Monitoring and Feedback: Establish a process for continuous monitoring and feedback based on sensitivity analysis results. Monitor changes in sensitive variables over time, validate model performance against observed data, and update forecasting methodologies iteratively to reflect evolving conditions and reduce forecast uncertainties.

Shapely method for building the sensitivity analysis.

The Shapley method is a statistical technique used to analyze the contribution of individual variables to the variability in a model's output. Applying this method to electricity load forecasting allows you to quantify the sensitivity of forecasts to changes in input variables. By identifying the most influential variables and their interactions, you can prioritize efforts to improve data quality or refine model assumptions, thereby reducing uncertainty in forecasts.

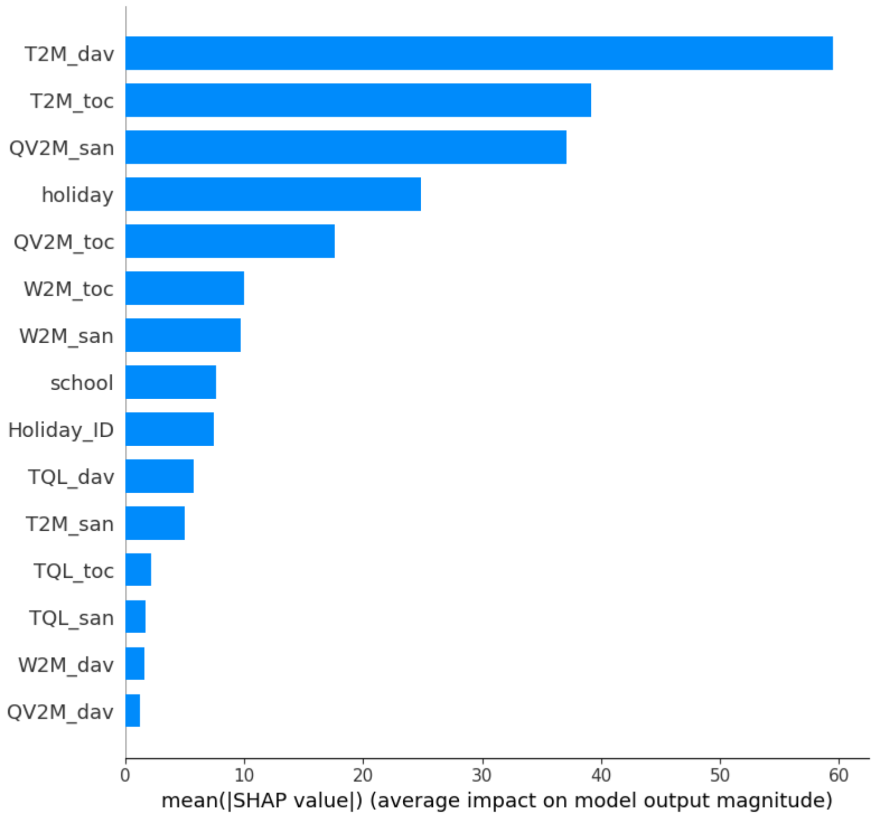
Regarding types of Shapley methods, there are variations and adaptations of the original Shapley value concept, tailored to specific applications and scenarios. Some types and adaptations include **Model-specific Shapley Values, Feature Importance Shapley Values, Ensemble Shapley Values, and Bootstrap Shapley Values.**

Methods used Model-specific Shapley Values and Temporal Shapley Values for this dataset.

Model-specific Shapley Values:

- This approach customizes the Shapley value computation to work directly with specific forecasting models or algorithms, leveraging the model's internal workings or feature importance measures.
- By incorporating model-specific information, such as coefficients in linear regression or feature importances in tree-based models, this method provides insights into how individual features contribute to model predictions.

In our Output:



Mean Squared Error (MSE):

The MSE value (18505.61) represents the average squared difference between the actual and predicted values of electricity load across all instances in the test set. A higher MSE indicates a larger discrepancy between the actual and predicted values, implying greater uncertainty in the model's predictions.

SHAP Summary Plot:

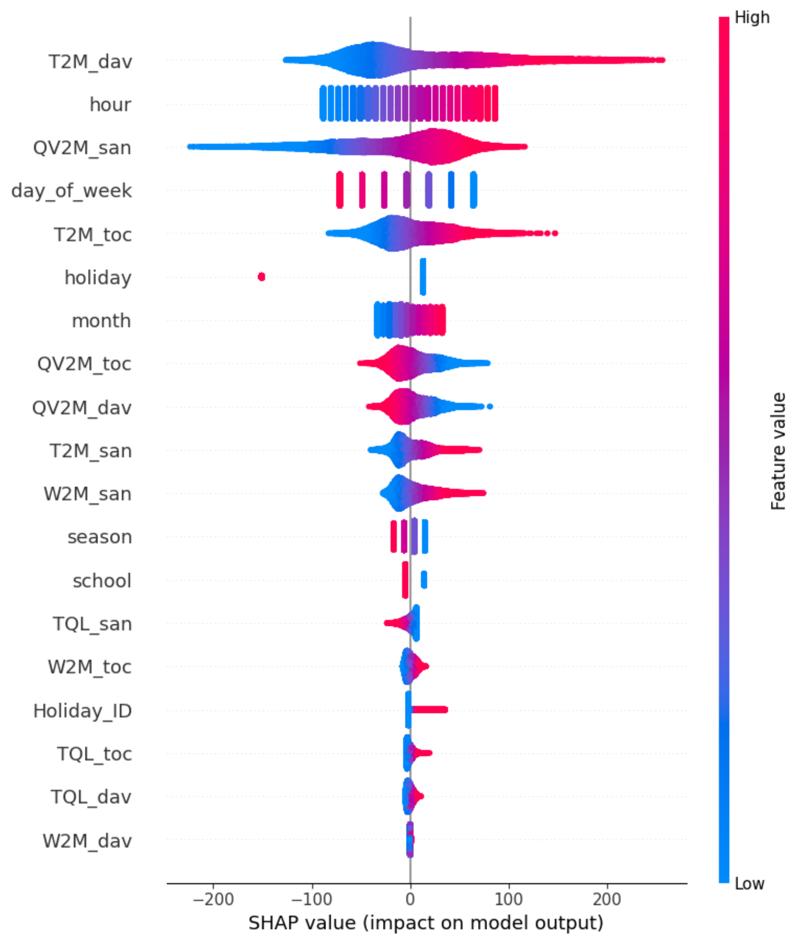
The SHAP summary plot displays the mean absolute SHAP values for each feature in the model. SHAP (SHapley Additive exPlanations) values represent the contribution of each feature to the difference between the actual prediction and the average prediction. Features with larger absolute SHAP values have a greater impact on the model's predictions.

Interpreting the SHAP Summary Plot:

T2M_dav: This feature has the highest mean absolute SHAP value, indicating that it has the most significant impact on the model's predictions. Specifically, an increase in T2M_dav leads to a decrease in electricity load.

T2M_toc, QV2M_san, holiday, QV2M_toc, W2M_toc, W2M_san, school, Holiday ID, TQL_dav, T2M_san, TQL_toc, TQL_san, W2M_dav, QV2M_dav: These features have lower mean absolute SHAP values compared to T2M_dav, indicating relatively lesser importance in predicting electricity load.

Overall, the SHAP summary plot provides insights into the importance of each feature in the Linear Regression model's predictions. By understanding the impact of different features, stakeholders can gain insights into the factors driving electricity load and how the model utilizes these factors to make predictions.



Mean Squared Error (MSE):

The MSE value (13991.11) represents the average squared difference between the actual and predicted values of electricity load across all instances in the test set. A higher MSE indicates a larger discrepancy between the actual and predicted values, implying greater uncertainty in the model's predictions.

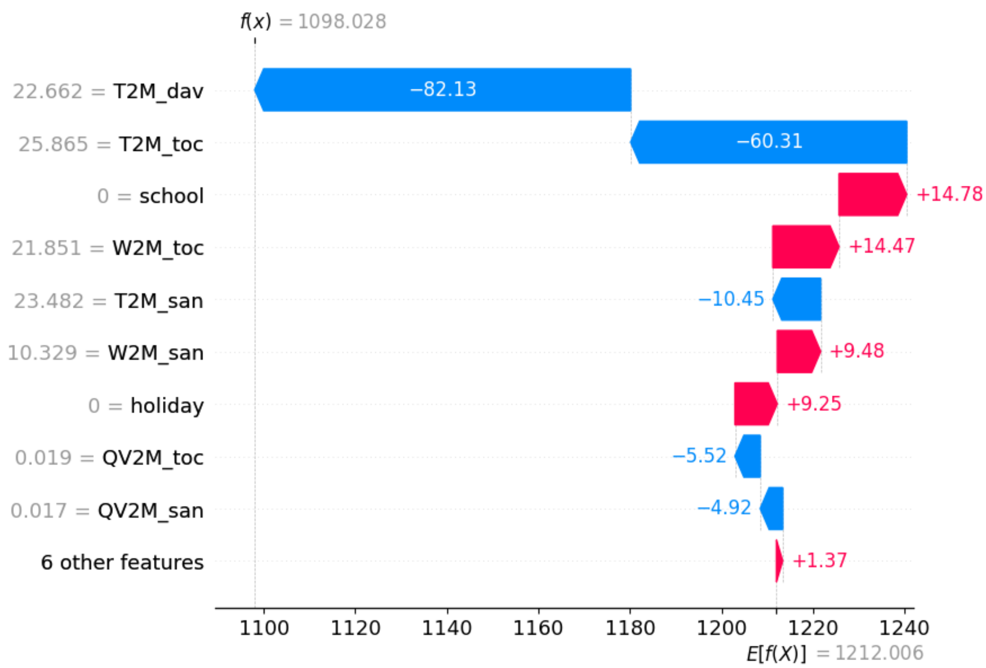
SHAP Summary Plot:

The SHAP summary plot displays the mean absolute SHAP values for each feature in the model. SHAP (SHapley Additive exPlanations) values represent the contribution of each feature to the difference between the actual prediction and the average prediction. Features with larger absolute SHAP values have a greater impact on the model's predictions.

Interpreting the SHAP Summary Plot:

T2M_dav: This feature has the highest mean absolute SHAP value, indicating that it has the most significant impact on the model's predictions. Specifically, an increase in T2M_dav leads to a decrease in electricity load.

hour, QV2M_san, day_of_week, T2M_toc, holiday, month, QV2M_toc, QV2M_dav, T2M_san, W2M_san, season, school, TQL_san, W2M_toc, Holiday ID, TQL_toc, TQL_dav, W2M_dav: These features have lower mean absolute SHAP values compared to T2M_dav, indicating relatively lesser importance in predicting electricity load.

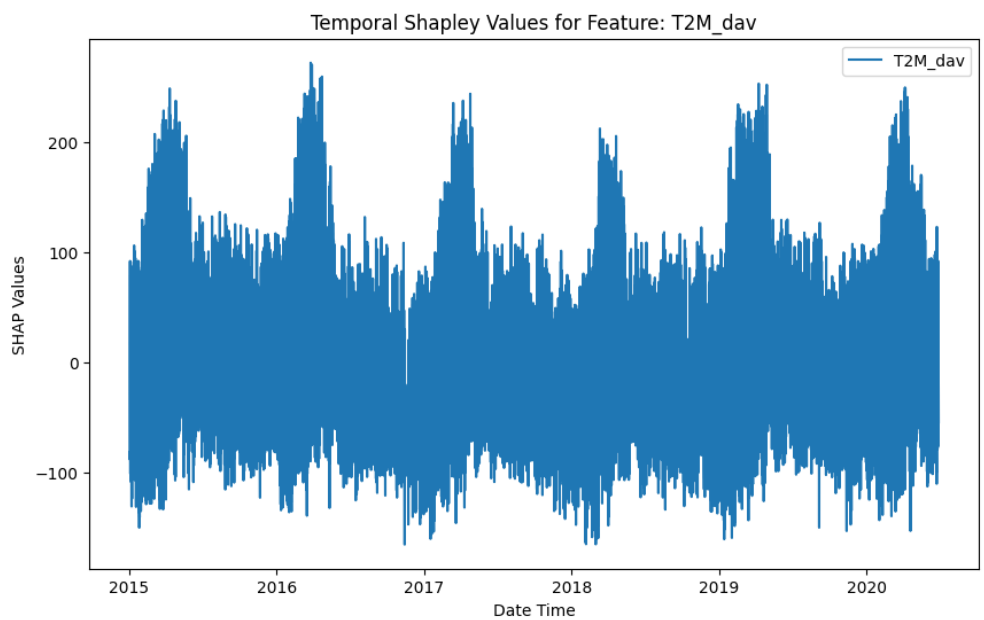


And this similarly t the above graphs shows the positive and negative effect of the features to the model that predicts the electrical load.

Temporal Shapley Values:

- Adapted to account for temporal patterns and dependencies in electricity load forecasting, this method considers how factors contribute to uncertainty at different time points or incorporates temporal dynamics into Shapley value computation.
- By analyzing the temporal dimension, stakeholders can gain insights into how variables impact forecasts over time, allowing for more nuanced risk assessment and mitigation strategies.

In our Output:



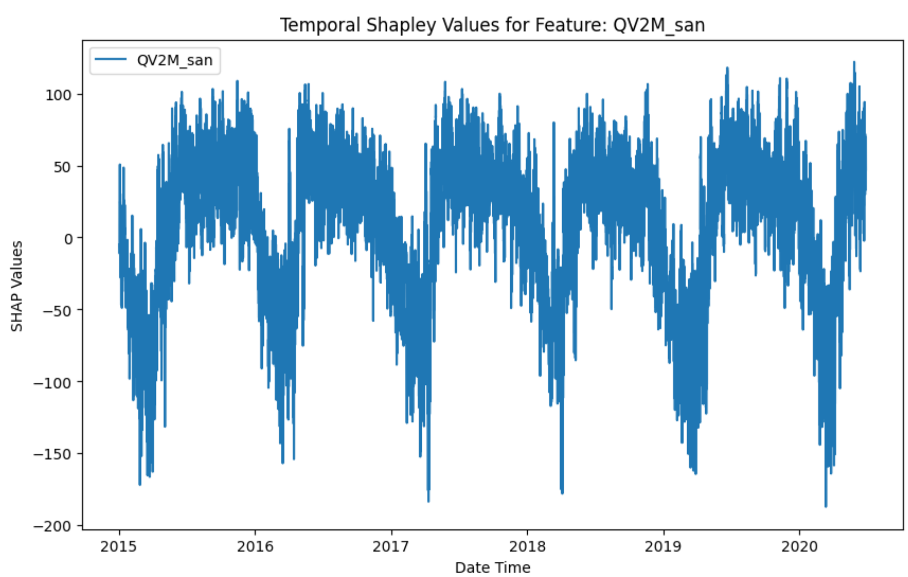
Features with higher absolute SHAP values indicate stronger influence on the model's predictions. In the graph, if we observe peaks or valleys for certain features at specific time points, it suggests that those features have a significant impact on the model's predictions during those periods.

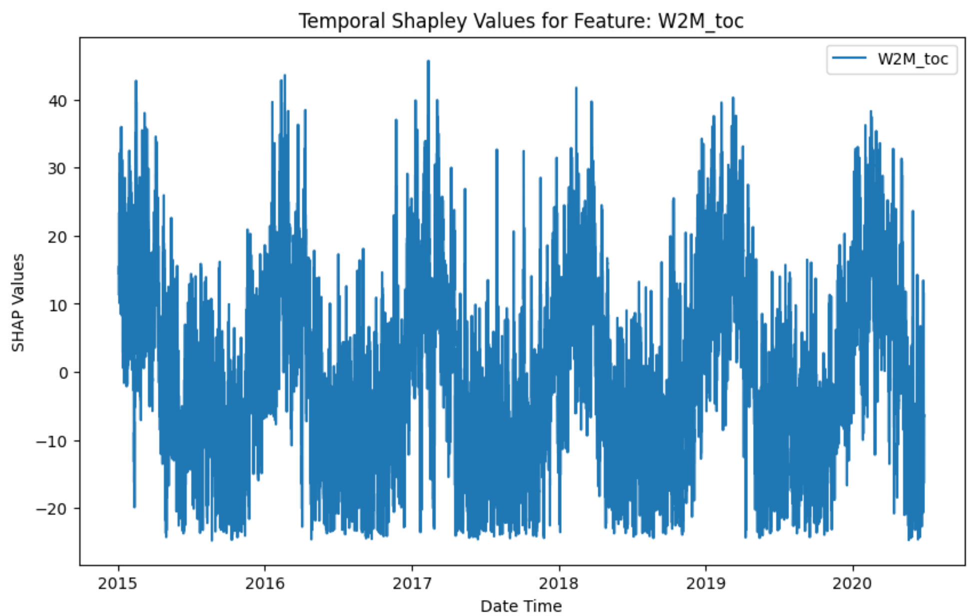
For example, if the SHAP values for the "T2M_dav" feature show consistently higher absolute values compared to other features across multiple time points, it suggests that "T2M_dav" has a strong influence on the model's predictions throughout the dataset's timeframe.

For instance, if we notice fluctuations in SHAP values for a particular feature over time, it indicates that the influence of that feature on the model's predictions varies across different periods. This temporal variation reflects the changing dynamics of the relationship between the feature and the target variable. Observing the graph, we can see that there is a fluctuation for "T2M_dav", but compared to the other features in is relatively not fluctuating as much.

The direction of SHAP values (positive or negative) at different time points provides insights into whether the feature positively or negatively contributes to the model's predictions during those periods. Hence, observing the direction of SHAP values, "T2M_dav" reaching 200 and falling to -100 signifies its alternating positive and negative impact on load forecasts, highlighting its dynamic nature.

The other features have similar observation as shown in the graphs below, there are others but this is to show some.





Simulation

Data simulation involves creating artificial data that closely resembles real-world data in terms of its properties and characteristics. Instead of gathering data through surveys, monitoring software, or web scraping, simulated data is generated using mathematical or computational models. This approach offers data scientists, engineers, and businesses access to training data at a reduced cost.

There are various types of data simulation models, each with its own unique features: **Monte Carlo simulations, Agent-based modeling, System dynamics and Discrete-event simulations.**

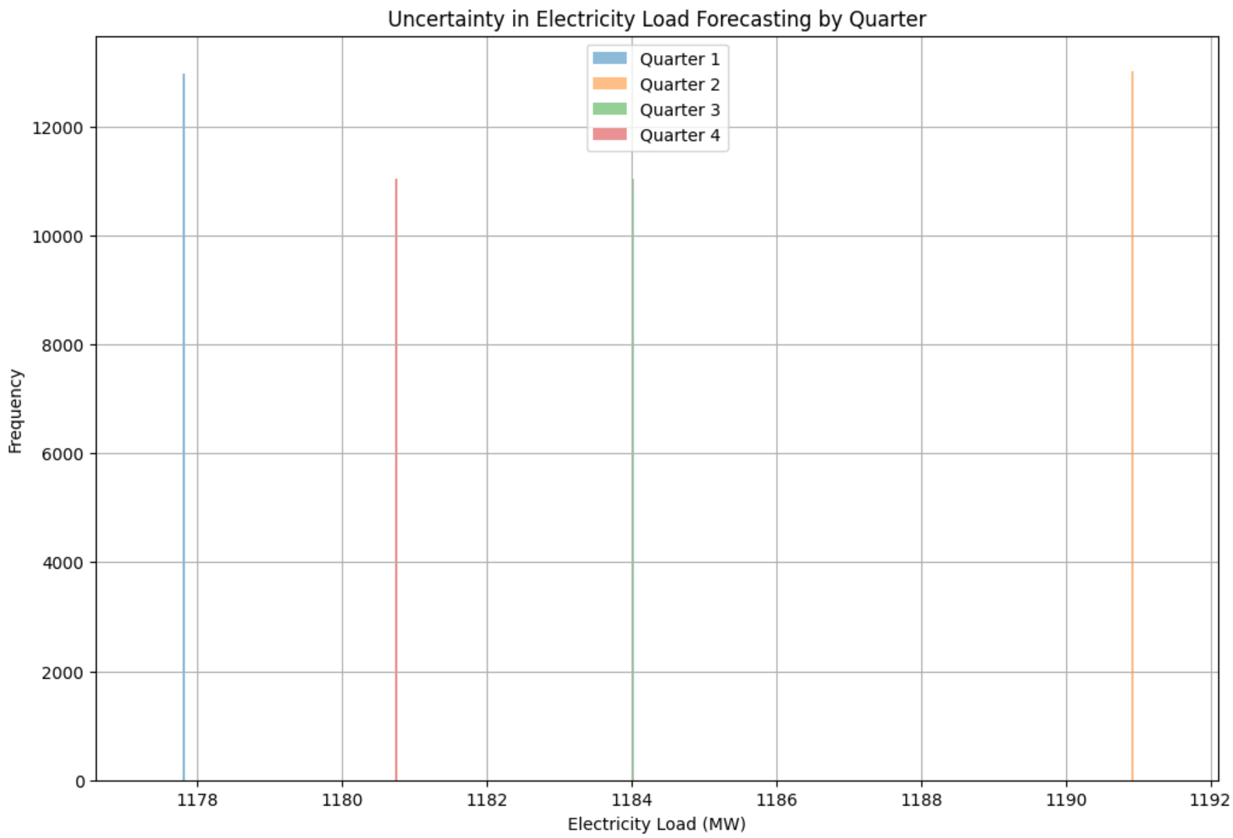
Monte Carlo simulations

- This method employs random sampling to provide insights into uncertain situations. It is commonly utilized in finance, physics, and engineering to model complex systems and forecast outcomes.

In our output:

A lower MAE indicates better model performance, as it signifies that, on average, the model's predictions are closer to the actual values. In this case, an MAE of 162.87 MW suggests that, on average, the model's predictions deviate from the actual values by approximately 162.87 MW.

Monte Carlo simulations are employed to visualize the uncertainty in electricity load forecasting for each quarter. Histograms are plotted, showing the distribution of forecasted electricity load demand for each quarter. The histograms provide insight into the range of possible outcomes and the associated frequencies. We can see that the quarters do not have the same electricity load and that the highest load is found in the 2nd quarter of the year. This can be due to the events which will be held in that quarter including Semana Santa (Holy Week) and the Festival del Manito Ocueño.



Discrete-event simulations

- These models concentrate on individual events within a system and their impact on the overall outcome. They are extensively used in operations research, computer science, and logistics to simulate processes and systems.

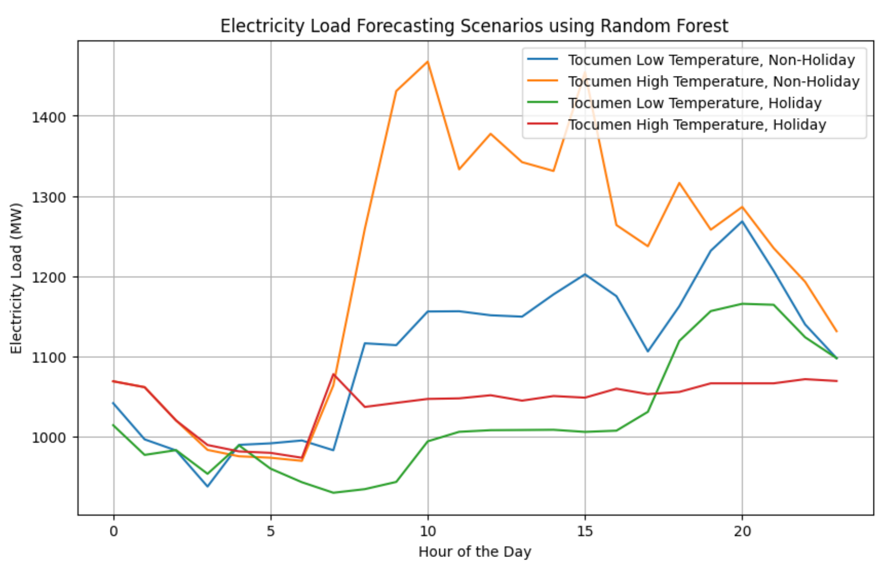
In our output:

The code visualizes the electricity load forecasts for each scenario using Random Forest.

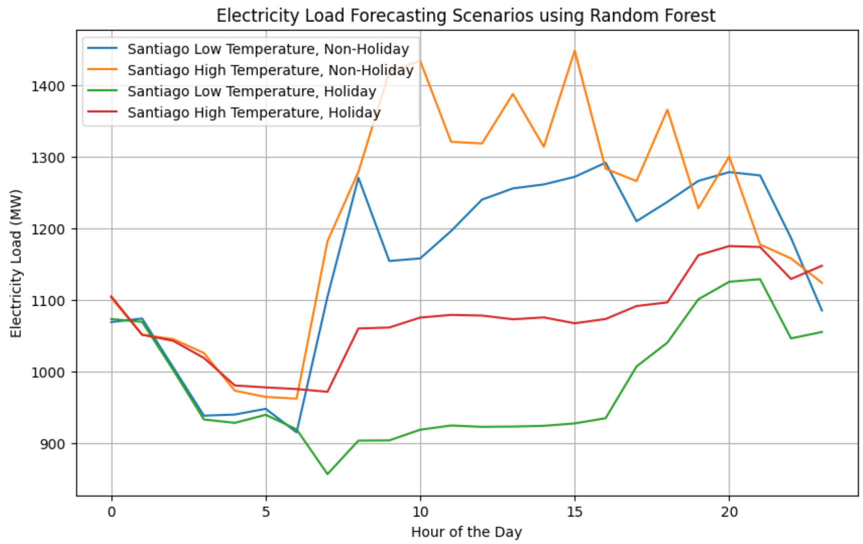
By plotting the forecasts for different scenarios, it becomes possible to observe how changes in temperature and holiday status affect the predicted electricity load throughout the day.

This analysis provides insights into the model's behavior under different conditions, helping to understand the impact of specific variables (temperature and holiday) on electricity load forecasting.

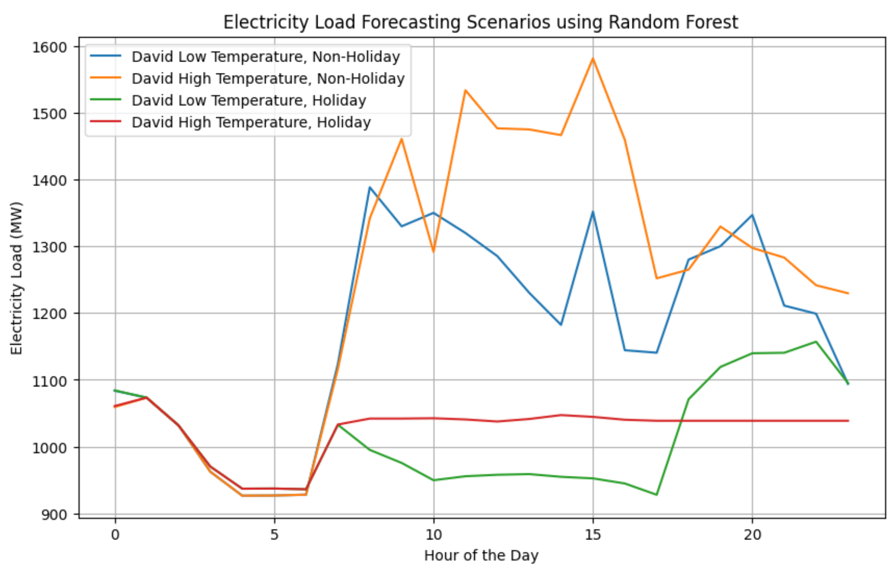
In this first graph we can see 4 scenarios for the city Tocumen, and in the second for Santiago and the third for David. This shows the effect of temperature and holiday in each city on electrical load.



In this graph we can see that in Tocumen there is high consumption of electricity load between the hour of 7 and 16 with 10 reaching its pick, when there is high temperature and no holidays. Which makes sense as people will be in their work and home turning the cooling system.



In this graph of Santiago, we see is similar to Tocumen in terms of high temperature and no holiday, but it is not as different from the rest of the scenarios as Tocumen.



In this last graph of David, we have similarity in term of high temperature and no holiday, but the low temperature and no holiday does also have high load compared with the ones that have holiday.

As a common factor in three of them between the hours 0 to about the electricity load if low which is the given as people could be sleeping. Additinaly, the scenarios with holiday are seen to have low electricity load as to compared the non-holiday scenarios.

Discontinuity regression

Discontinuity regression is a statistical method employed to analyze data exhibiting abrupt changes or discontinuities in the relationship between independent and dependent variables. Unlike traditional regression techniques that assume a smooth and continuous relationship, discontinuity regression acknowledges and explicitly models these abrupt shifts. This approach is particularly valuable in scenarios where specific events, thresholds, or interventions lead to sharp changes in the outcome of interest. By identifying and estimating the effects of such discontinuities, discontinuity regression provides insights into causal relationships and helps evaluate the impact of interventions or policies.

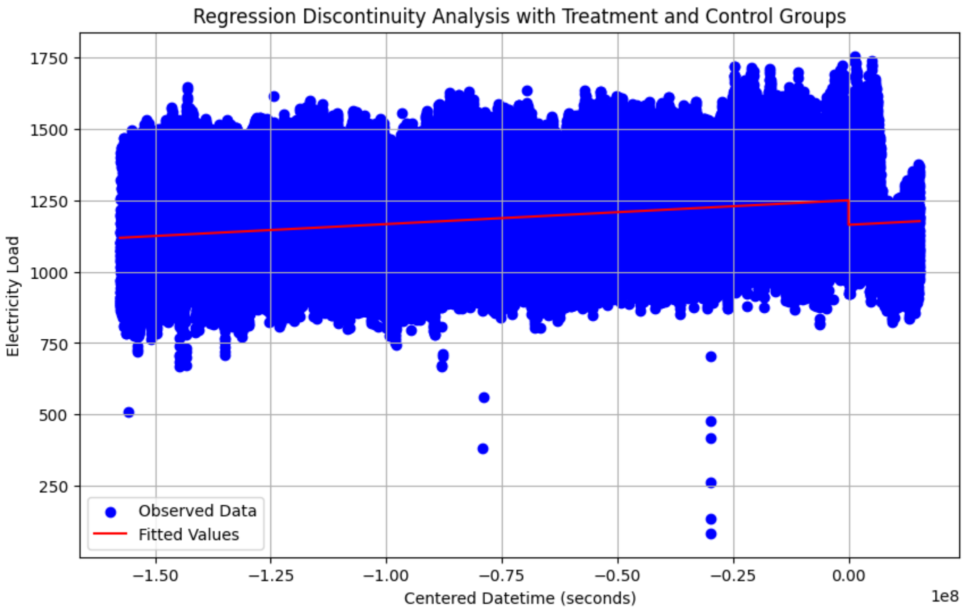
Types of Discontinuity Regression Models: **Regression Discontinuity Design (RDD), Sharp Regression Discontinuity Design (SRDD), Fuzzy Regression Discontinuity Design (FRDD), Local Regression Discontinuity Design (LRDD), Discontinuity-in-Density Design (DiD), and Segmented Regression Analysis.**

Regression Discontinuity Design (RDD):

- RDD compares observations around a predefined threshold or discontinuity point.
- It estimates the treatment effect precisely at the threshold by focusing on observations just above and below it.

In our output:

In terms of Datetime



Interpretation of Model Results:

The RDD analysis helps identify whether there is a systematic change in electricity load around a specific cutoff datetime (e.g., "2020-01-01" in the code) and whether this change is attributable to the datetime variable. The cutoff date time was chosen to see the effect of the start of COVID-19.

The regression results provide insights into the relationship between centered datetime, treatment group, and electricity load.

The coefficient for centered datetime (8.326e-07) indicates the average change in electricity load for a one-unit increase in centered datetime, holding the treatment group constant. The statistical significance (p-value < 0.001) suggests that there is a significant association between datetime and electricity load.

The coefficient for the treatment group (-85.8392) represents the discontinuous change in electricity load between the treatment and control groups at the cutoff datetime. The negative coefficient indicates that the treatment group (datetime > 0) experiences a decrease in electricity load compared to the control group (datetime <= 0).

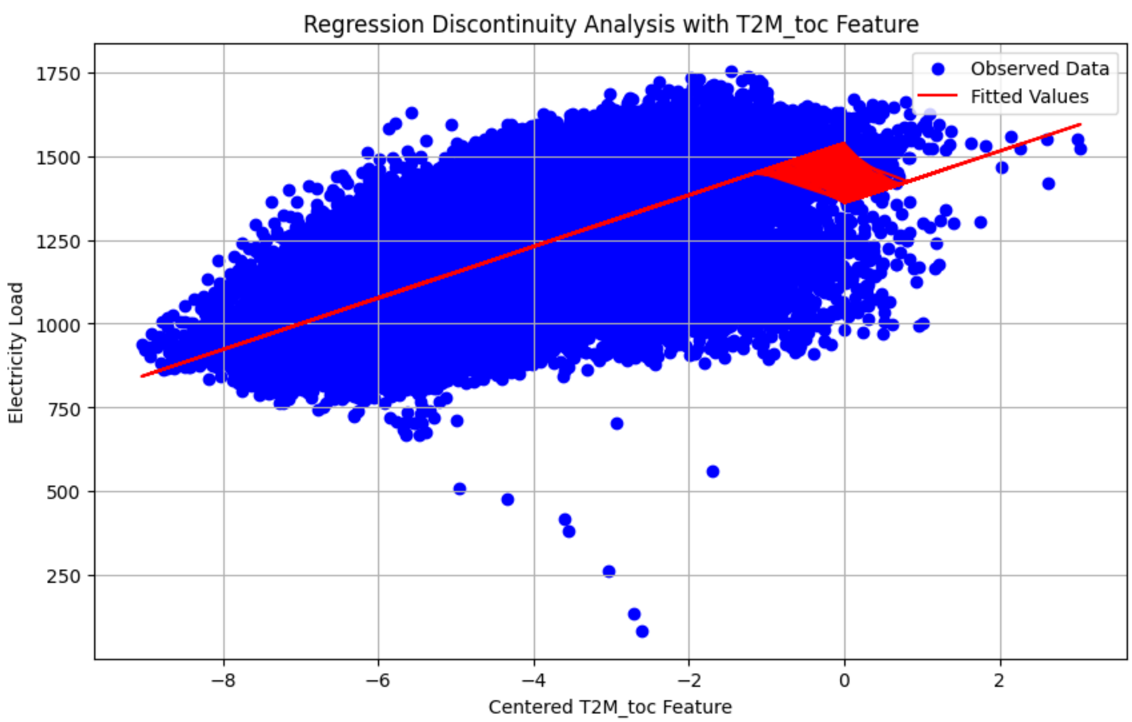
The R-squared value (0.036) indicates the proportion of variance in electricity load explained by the predictors in the model. While relatively low, this suggests that datetime and treatment group collectively explain a small but statistically significant portion of the variability in electricity load.

Implications for Mitigating Risks:

The RDD analysis helps identify systematic changes in electricity load around the cutoff datetime, providing insights into potential external factors or events influencing load patterns.

By understanding the relationship between datetime and electricity load, stakeholders can develop strategies to mitigate risks associated with forecast uncertainty. For example, if the analysis reveals a significant decrease in load after a specific datetime, stakeholders can investigate the underlying causes (e.g., changes in weather patterns, policy interventions) and adjust forecasting models accordingly.

In terms of T2M_toc



RDD is a quasi-experimental research design used to estimate causal effects by exploiting a sharp change or discontinuity in a treatment variable. In this case, the treatment variable is the centered feature (e.g., T2M_toc) around a cutoff value which is the temperature 32 in our graph.

Interpretation of Model Results:

The regression results provide insights into the relationship between the centered feature, treatment group, and electricity load.

The coefficient for the centered feature (76.7980) indicates the average change in electricity load for a one-unit increase in the centered feature, holding the treatment group constant. The statistical significance (p-value < 0.001) suggests that there is a significant association between the alternative feature and electricity load.

The coefficient for the treatment group (-176.2490) represents the discontinuous change in electricity load between the treatment and control groups at the cutoff value of the alternative feature. The negative coefficient indicates that the treatment group (centered feature > 0) experiences a decrease in electricity load compared to the control group (centered feature <= 0).

The R-squared value (0.431) indicates the proportion of variance in electricity load explained by the predictors in the model. This relatively higher value compared to the previous example suggests that the alternative feature (T2M_toc) has a stronger explanatory power for electricity load variability.

Implications for Mitigating Risks:

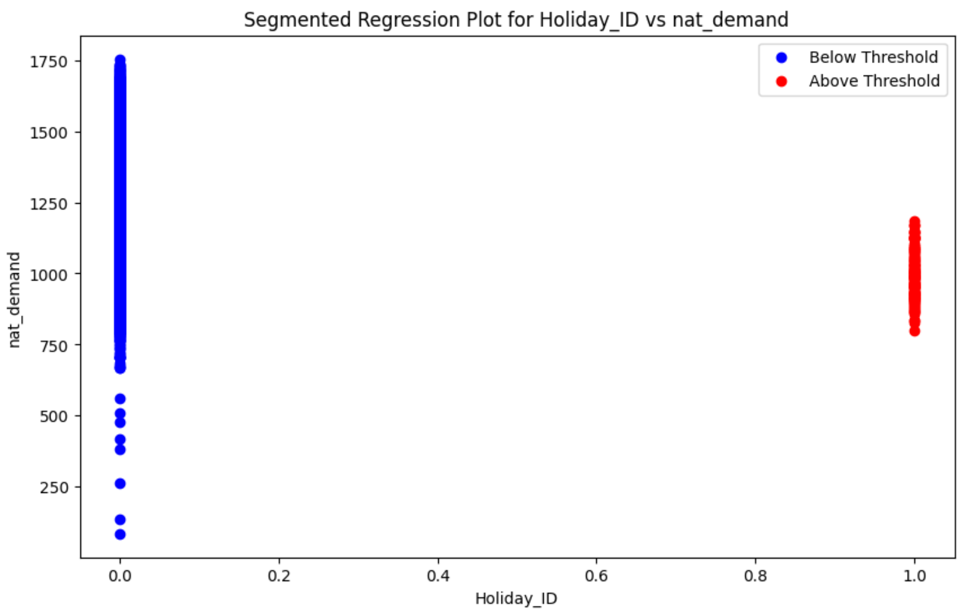
The RDD analysis helps identify systematic changes in electricity load around the cutoff value of the alternative feature, providing insights into potential external factors or events influencing load patterns.

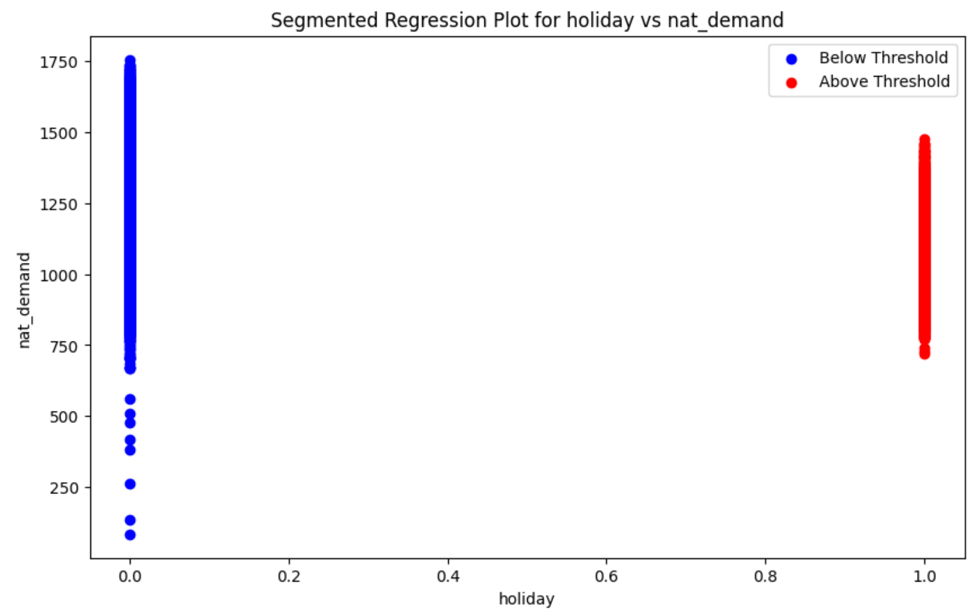
By understanding the relationship between the alternative feature and electricity load, stakeholders can develop strategies to mitigate risks associated with forecast uncertainty. For example, if the analysis reveals a significant decrease in load beyond a certain temperature threshold, stakeholders can incorporate this information into their forecasting models to improve accuracy and reliability.

In summary, the RDD analysis conducted in the provided code contributes to addressing question 3 by elucidating the relationship between an alternative feature (e.g., T2M_toc) and electricity load, thereby informing strategies to mitigate risks associated with forecast uncertainty in electricity load forecasting.

Segmented Regression Analysis:

- Segmented Regression Analysis models relationships between variables when the relationship isn't linear throughout.
- It fits separate regression lines to different data segments, allowing for discontinuities or non-linearities in the relationship. This provides a more flexible and accurate representation of complex data relationships.





The code iterates through each threshold (Holiday_ID, holiday) and splits the dataset into two groups based on these thresholds: below the threshold and above the threshold.

For each group, it fits a segmented regression model to estimate the relationship between the feature (threshold) and the electricity load.

The segmented regression allows for different regression lines on either side of the threshold, capturing any discontinuities or changes in the relationship between the feature and the target variable.

Interpretation of Segmented Regression Plots:

The vertical lines in the plot represent the thresholds (Holiday_ID = 0 or 1, holiday = 0 or 1), indicating the points where the regression lines may have discontinuities or changes in slope.

Observing the discontinuities or changes in the regression lines at the thresholds suggests potential impacts of these features on electricity load. For example, the abrupt changes in electricity load at the threshold points may indicate the influence of holidays or specific events on electricity demand.

The range of electricity load values (e.g., from 0 to 1750 for Holiday_ID = 0 and from 800 to 1150 for Holiday_ID = 1) provides insights into the variability in load demand under different conditions or scenarios, contributing to understanding uncertainty in load forecasting.

Implications for Mitigating Risks:

By identifying the discontinuities or changes in the relationship between features (e.g., holidays) and electricity load, stakeholders can better understand the factors contributing to uncertainty in load forecasting.

This analysis can inform strategies to mitigate risks associated with forecast uncertainty, such as adjusting resource allocation, optimizing energy production and distribution, or implementing demand-side management measures during critical periods identified by the segmented regression analysis.

Explainability of variables

The explainability of variables refers to the understanding of how individual input features or variables contribute to the output or outcome of a model. In many machine learning and statistical modeling tasks, it's crucial to interpret and explain the influence of different features on the model's predictions or outcomes. By understanding the explainability of variables, analysts, data scientists, and stakeholders gain insights into the underlying relationships between input features and the target variable. This understanding can inform decision-making processes, enhance model transparency, and identify influential factors driving the model's behavior.

Types of Techniques for Explaining Variable Influence: **Feature Importance, Partial Dependence Plots (PDPs), Individual Conditional Expectation (ICE) Plots, LIME (Local Interpretable Model-agnostic Explanations), SHAP Values (SHapley Additive exPlanations), and Model-Specific Interpretations.**

Used method with analysis.

SHAP Values (SHapley Additive exPlanations):

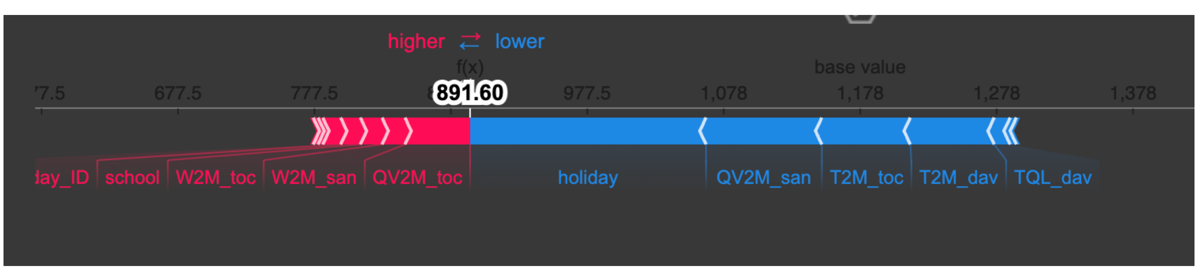
SHAP values provide global or local explanations by attributing each feature's impact on the model's output to its Shapley value, which reflects the feature's contribution to the average prediction.

Mean Squared Error (MSE):

The MSE value (18505.61) represents the average squared difference between the actual and predicted values of electricity load across all instances in the test set. A higher MSE indicates a larger discrepancy between the actual and predicted values, implying greater uncertainty in the model's predictions. SHAP Summary Plot:

The SHAP summary plot visualizes the average impact of each feature on the model's output across all instances in the dataset. Features are ranked based on their importance in explaining the output variability. Positive SHAP values indicate a positive impact on the prediction, while negative values indicate a negative impact. Incorporating Domain Knowledge:

The code snippet calculates the average SHAP values for specific temperature-related variables (T2M_toc, QV2M_san, T2M_dav). These variables are likely related to weather conditions, and their SHAP values can help interpret their significance in influencing electricity demand. For example, a positive average SHAP value for temperature (T2M_toc) suggests that higher temperatures are associated with increased electricity demand, while a negative value for humidity (QV2M_san) implies a negative impact on demand.



The force plot that visualizes the contributions of each feature to the model's prediction for that instance. Each feature's contribution is represented by a colored bar, where positive values indicate an increase in the predicted load, and negative values indicate a decrease. The length of the bar represents the magnitude of the contribution. Overall, SHAP values provide insights into how each feature contributes to individual predictions, helping to interpret model decisions and understand the factors influencing electricity load forecasting.

LIME (Local Interpretable Model-agnostic Explanations):

LIME generates local explanations for individual predictions by approximating the model's behavior in the vicinity of a data point using interpretable surrogate models.

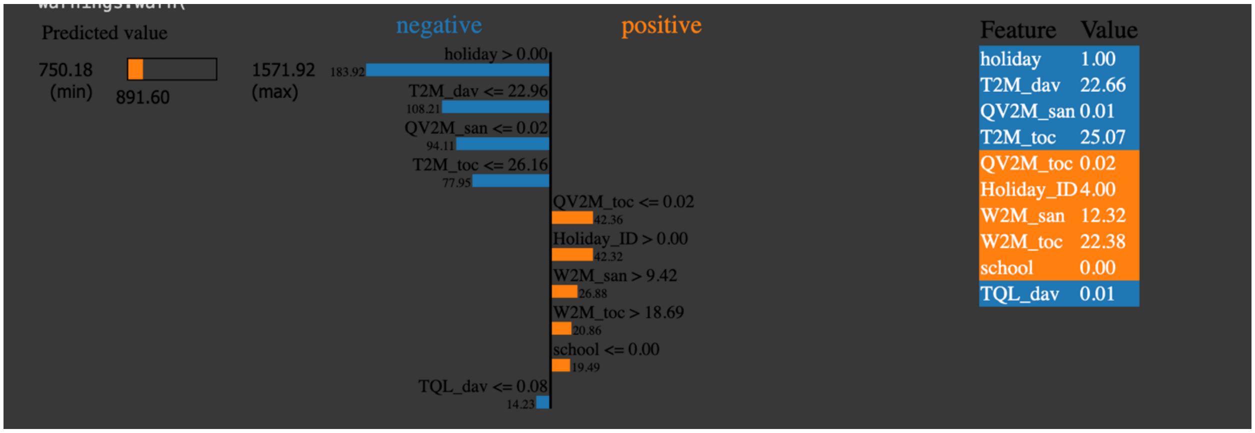
The provided output is generated by LIME (Local Interpretable Model-agnostic Explanations) to explain the prediction made by a linear regression model for a specific instance of electricity load forecasting.

Mean Squared Error (MSE):

The MSE value (18505.61) represents the average squared difference between the actual and predicted values of electricity load across all instances in the test set. A higher MSE indicates a larger discrepancy between the actual and predicted values, implying greater uncertainty in the model's predictions. Instance Features:

The instance features represent the input variables used to predict the electricity load for a specific time-period. These features include various weather-related variables (temperature, humidity), holiday indicators, and other contextual factors.

The predicted value (891.60) is the model's forecasted electricity load for the given instance. It represents the central estimate or point prediction made by the linear regression model.

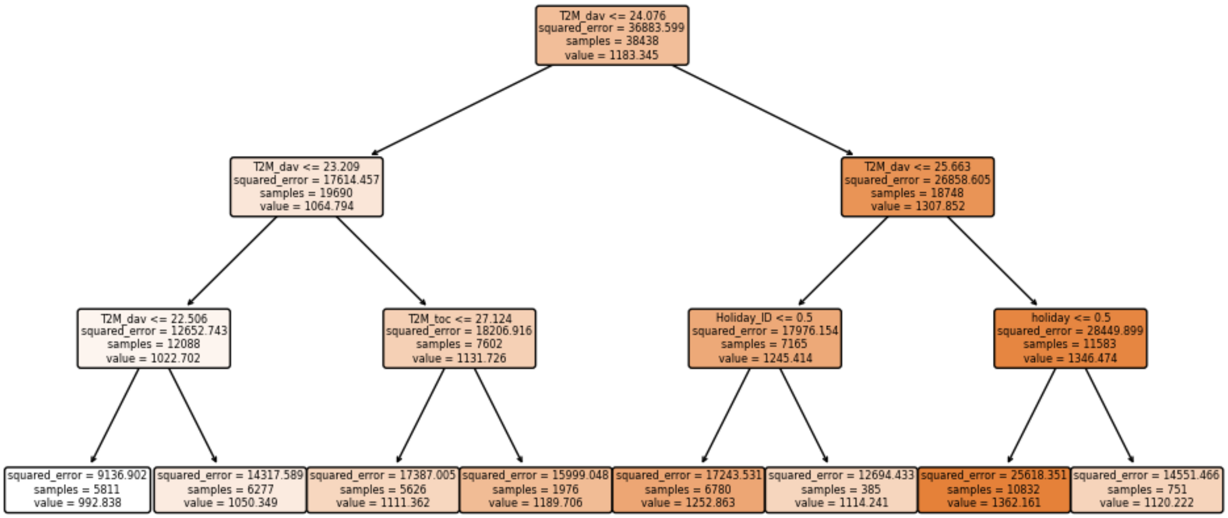


LIME provides an uncertainty range by computing the minimum and maximum predicted values across multiple perturbed instances. This range indicates the variability or uncertainty in the model's predictions for the given instance. In this case, the uncertainty range spans from 750.18 to 1571.92. A wider range suggests higher uncertainty in the prediction. Feature Contributions:

LIME also highlights the contributions of individual features to the predicted electricity load. It shows which features have the most significant impact on the prediction. Features are ranked based on their contribution to the predicted value. For example, the presence of a holiday (holiday > 0.00) and specific weather conditions (T2M_dav <= 22.96) are associated with higher or lower electricity load predictions.

For each feature, LIME provides the corresponding value for the given instance. This allows you to understand the context in which the feature is influencing the prediction. For instance, if the holiday feature has a value of 1.00, it indicates that the prediction was made for a holiday period. Overall, this output from LIME helps in interpreting the linear regression model's prediction for a specific instance of electricity load forecasting, including understanding the central estimate, uncertainty range, and the impact of individual features on the prediction.

Feature importance and Decision Tree.



The decision tree assigns a significant importance (approximately 92.23%) to the "T2M_dav" feature, which likely represents the temperature at a certain location.

The "Holiday_ID" feature also contributes to the predictions with an importance of around 6.54%. This indicates that whether a day is a holiday or not influences electricity demand, according to the decision tree.

Other features like "T2M_toc" (temperature at another location) have relatively low importance, around 1.22%, suggesting they have less impact on the model's predictions compared to "T2M_dav" and "Holiday_ID".

Features such as "QV2M_toc", "TQL_toc", "W2M_toc", "T2M_san", "QV2M_san", "TQL_san", "W2M_san", "QV2M_dav", "TQL_dav", "W2M_dav", "holiday", and "school" have no importance according to the decision tree, indicating that they might not significantly affect the model's predictions based on the tree's splits.

This interpretation provides insights into which features the decision tree considers most important for predicting electricity demand, highlighting the role of temperature and holidays while discounting the influence of other features in this particular model.

Mitigating risk:

The output allows stakeholders to understand how different features contribute to the model's predictions. By identifying the most important features, stakeholders can focus on collecting high-quality data and refining model assumptions related to those features. For instance, if "T2M_dav" is the most important feature, stakeholders might prioritize improving the accuracy of temperature measurements or incorporating additional weather-related variables to enhance forecasting accuracy. Ultimately, this helps reduce uncertainty in electricity load forecasts and improve decision-making processes.

Conclusion

The methods employed in electricity load forecasting offer a comprehensive approach to mitigating uncertainty and improving forecast accuracy. Techniques such as uncertainty analysis through bootstrapping provide a means to quantify variability in load predictions, while scenario analysis enables stakeholders to assess a range of potential outcomes and plan accordingly. Sensitivity analysis and Shapley methods identify influential factors driving load variations, guiding efforts to refine models and reduce uncertainty. Temporal Shapley values further enhance understanding by revealing how feature importance evolves over time, aiding in the anticipation of future uncertainties. Simulation methods like Monte Carlo and discrete-event simulations simulate diverse scenarios, enabling exploration of a wide range of potential outcomes and their associated uncertainties. Discontinuity regression and RDD techniques help identify abrupt changes in load behavior, while segmented regression analysis captures variations across different load patterns. Together, these methodologies provide valuable insights and tools for decision-makers in energy systems, ultimately enhancing the reliability and effectiveness of electricity load forecasting amidst uncertainty.

Some general mitigation strategies:

➢ Ensemble Methods: Combine multiple simulation runs or ensemble different forecasting models to reduce the impact of uncertainties inherent in individual predictions.

- ➢ Utilize ensemble techniques like bagging or boosting to aggregate predictions from multiple trees and mitigate the risk of overfitting.
- ➢ Random Forest with Simulation Approach: Increase the number of simulation scenarios to capture a broader range of possible outcomes and better quantify uncertainty.
- ➢ Monte Carlo Simulations: Generate a larger number of Monte Carlo samples to reduce sampling error and provide a more accurate estimation of uncertainty bounds.
- ➢ Regularly validate model performance using appropriate metrics, such as mean absolute error or mean squared error, on holdout or out-of-sample datasets to ensure robustness and reliability of forecasts.

Understanding Seasonal Patterns in Electricity Consumption

FB PROPHET:

Introduction

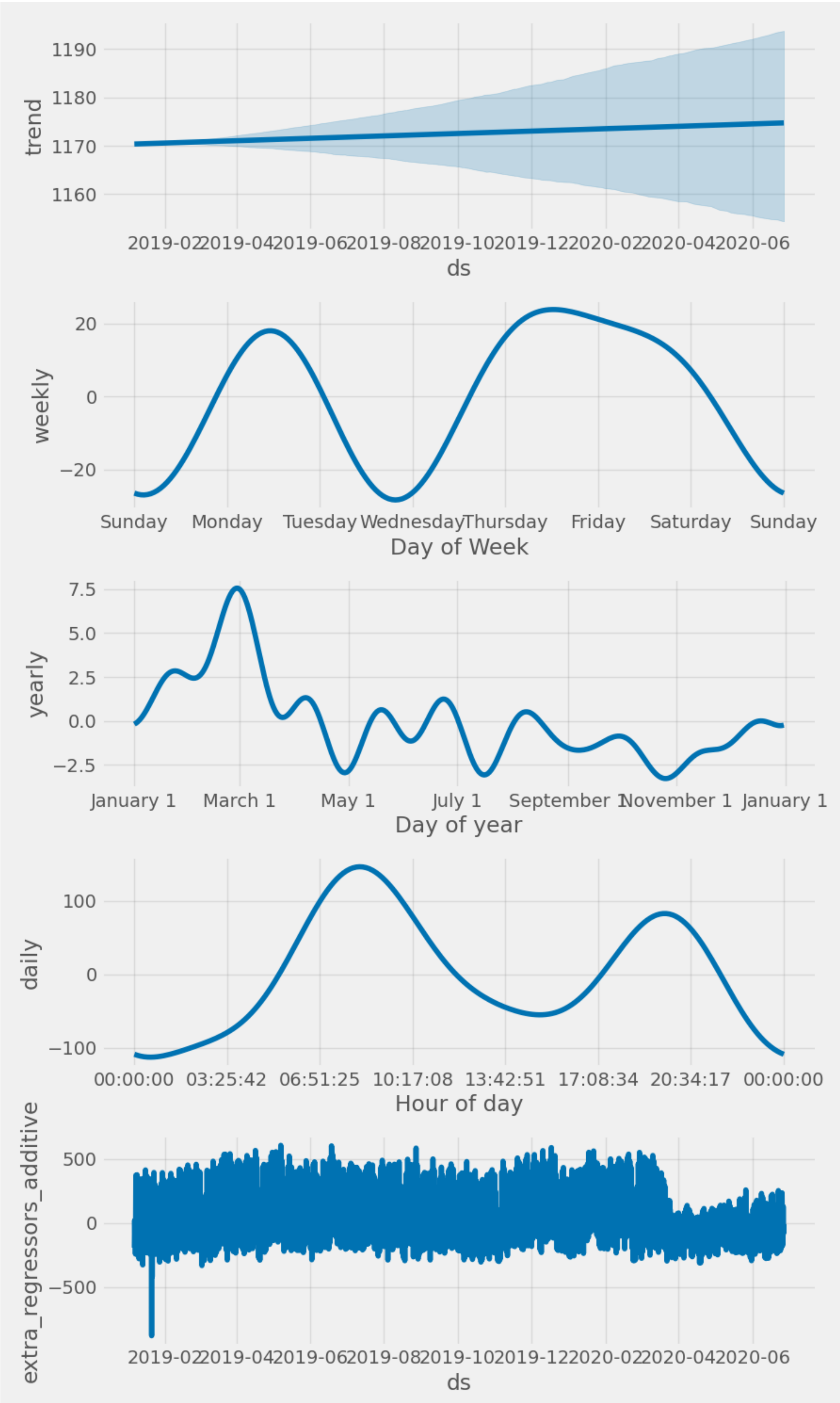
Facebook Prophet is a robust tool for time series forecasting, particularly effective in capturing seasonal patterns and trends in data. In this project, we leverage Facebook Prophet to analyze historical electricity load data, aiming to identify seasonal patterns and variations in consumption over different times, such as seasons or months.

Advantages of Facebook Prophet

Facebook Prophet's ability to handle various seasonalities, including daily, weekly, and yearly patterns, makes it ideal for analyzing electricity consumption data. Its user-friendly design, requiring minimal parameter tuning, makes it accessible to users without extensive time series analysis experience.

Analysis Overview

We utilized the Facebook Prophet model to analyze historical electricity load data, identifying seasonal patterns, holiday effects, and long-term consumption trends. The goal was to understand how electricity consumption changes over time and how the model can make accurate predictions. We evaluated the model's performance using metrics such as R-squared, RMSE, and MAE, and examined the trend analysis for insights into long-term consumption patterns.


trend
1190
1180
1170
1160
2019-02
2019-04
2019-06
2019-08
2019-10
2019-12
2020-02
2020-04
2020-06
ds
weekly
20
0
−20
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week
yearly
7.5
5.0
2.5
0.0
−2.5
January 1
March 1
May 1
July 1
September 1
November 1
January 1
Day of year
daily
100
0
−100
00:00:00
03:25:42
06:51:25
10:17:08
13:42:51
17:08:34
20:34:17
00:00:00
Hour of day
extra_regressors_additive
500
0
−500
2019-02
2019-04
2019-06
2019-08
2019-10
2019-12
2020-02
2020-04
2020-06
ds

Graph Explanation

- Trend: The trend analysis of the Facebook Prophet model reveals a consistent and noticeable increase in electricity consumption over time, suggesting a growing demand.
- Weekly Pattern: The weekly pattern shows higher demand at the start of the week, decreasing towards the middle, and peaking again on Thursday, gradually decreasing through the weekend.
- Daily Pattern: There is generally high demand between 6:51 am and 10:51 am, gradually decreasing and peaking slightly in the evening around 6:30 pm to 9:00 pm.
- Yearly Pattern: From January to mid-April, there is a high demand, possibly due to weather conditions requiring heaters.
- Extra Regressor Additive: This graph shows the influence of other features in the dataset on demand.

Seasonal Patterns

The Facebook Prophet model effectively captures seasonal patterns in electricity consumption, with an R-squared value of approximately 0.95, indicating a strong correlation. For example, the model detects higher consumption during certain seasons, such as summer or winter, reflecting increased demand for cooling or heating.

Holiday Effects

Incorporating the impact of holidays on electricity consumption, the model provides more accurate predictions during holiday periods. This is evident in the low RMSE and MAE values, indicating close alignment between predicted and actual values.

Trend Analysis

Facebook Prophet's trend analysis reveals a gradual increase in electricity consumption over several years, suggesting a growing demand. This insight is valuable for energy providers in making informed decisions about infrastructure investments and capacity planning.

Forecasting Accuracy

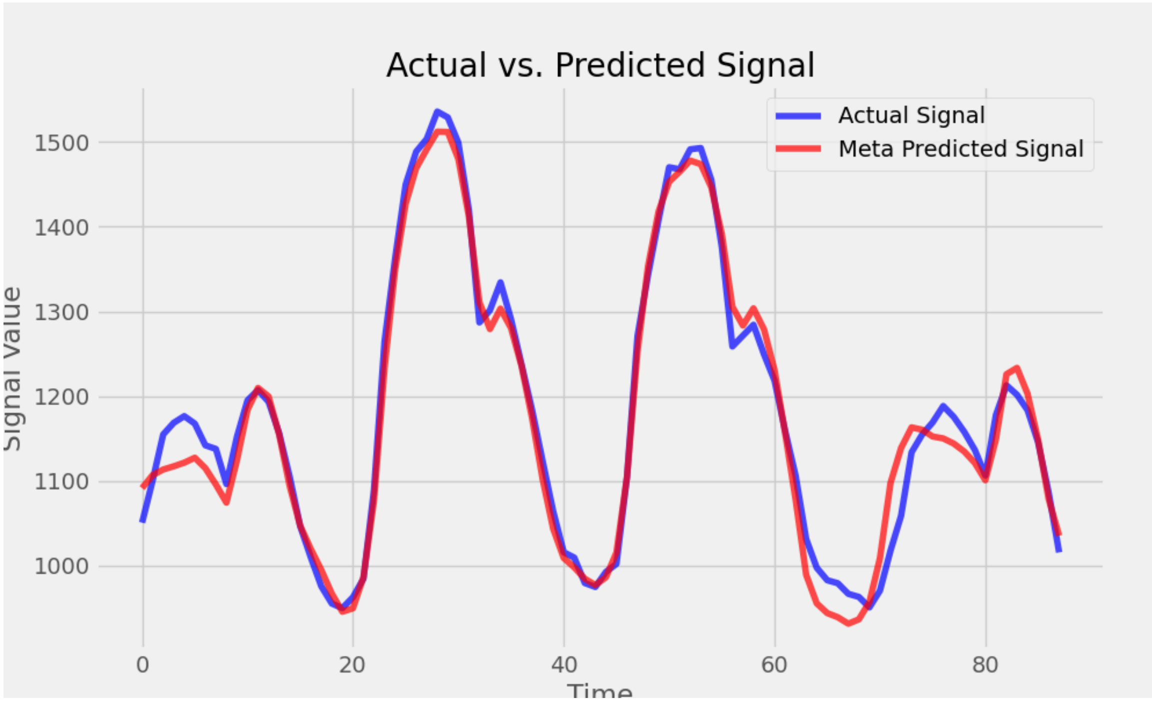
The model demonstrates high forecasting accuracy, with low RMSE and MAE values (approximately 41.83 and 30.02, respectively), indicating close alignment between predicted and actual consumption values. The R-squared value of approximately 0.95 further confirms the model's precision in predicting electricity consumption.

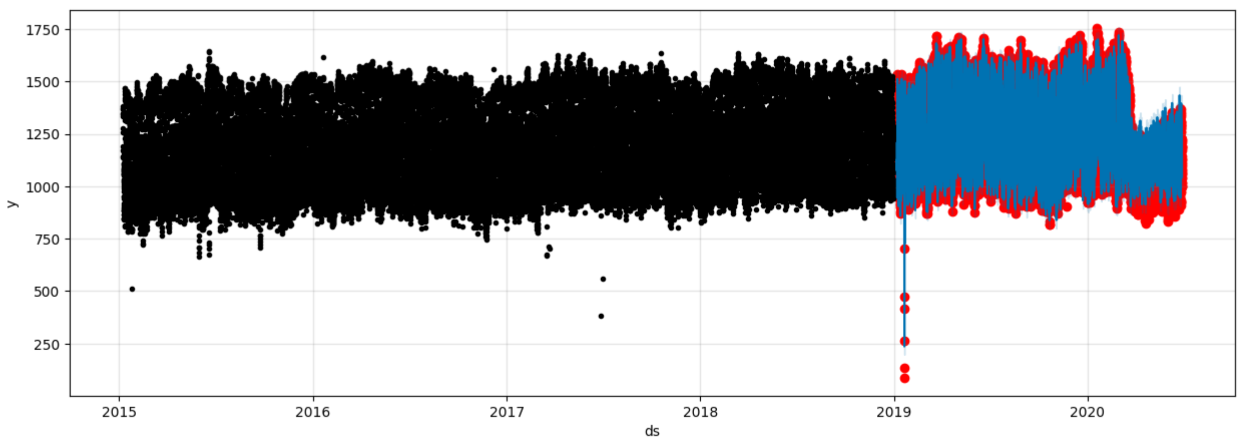
Model Performance

With low Mean Poisson Deviance and Mean Gamma Deviance values (approximately 1.62 and 0.0017, respectively), the model fits well to the data distribution. By accurately capturing consumption data distribution, the model provides valuable insights for energy management and planning.
Summary of evaluation metrics:

Metric	Value
RMSE	41.83
MAE	30.02
MAPE	0.0267
R-squared	0.9505
Explained Variance	0.9543
Max Error	622.03
Mean Poisson Deviance	1.6151
Mean Gamma Deviance	0.0017
Mean Tweedie Deviance	1749.92

Predicted vs. Actual





Another important aspect of the analysis is comparing the model's predictions with the actual consumption values. The table below shows a sample of the predicted values generated by the Facebook Prophet model compared to the actual consumption values. The model's predictions are generally close to the actual values, indicating its effectiveness in capturing the underlying patterns in the data.

Date	Actual	Prophet Prediction
2019-01-06 09:00:00	**1050.7971**	**1091.582309**
2019-01-06 10:00:00	**1100.5710**	**1106.937200**
2019-01-06 11:00:00	**1154.8653**	**1113.502435**
2019-01-06 12:00:00	**1168.6314**	**1117.293967**
...	**...**	**...**

Graph of actual versus predicted

To summarize the analysis using Facebook Prophet for understanding seasonal patterns in electricity consumption provides valuable insights for energy management and planning. The model's ability to capture seasonal variations, holiday effects, and long-term trends makes it a valuable tool for energy providers to optimize resources and meet varying demand throughout the year.From the output results and the provided table, we can see that the Prophet model's predictions are generally close to the actual values. These metrics suggest that the Prophet model is effective in forecasting electricity demand, with relatively small errors in its predictions. It appears that there is a repetitive pattern in the dataset as it can be seen in the trends, weekly, yearly and daily graphs of the model. The model seems to capture the underlying trends and seasonality in the electricity demand data, as evidenced by the close alignment between the actual and predicted values. The repetition in the pattern likely reflects the daily and possibly weekly or monthly fluctuations in electricity demand, which the model is able to capture and forecast not as accurately but acceptable.

Autoregressive Model Analysis for Electricity Demand Forecasting

Autoregressive (AR) models play a crucial role in forecasting electricity demand based on historical consumption patterns. These models examine how past electricity load values relate to future load values, helping to identify and predict recurring patterns such as daily, weekly, or seasonal variations in consumption. Understanding these patterns is essential for effective energy management and planning.

How Autoregressive Models Work

Autoregressive models analyze the relationship between past and future electricity load values without relying on complex mathematical formulas. Instead, they look for patterns in the data by assessing how changes in past load values correspond to changes in future load values. By recognizing these patterns, such as higher consumption during certain times of the day, week, or year, AR models can make informed predictions about future load levels.

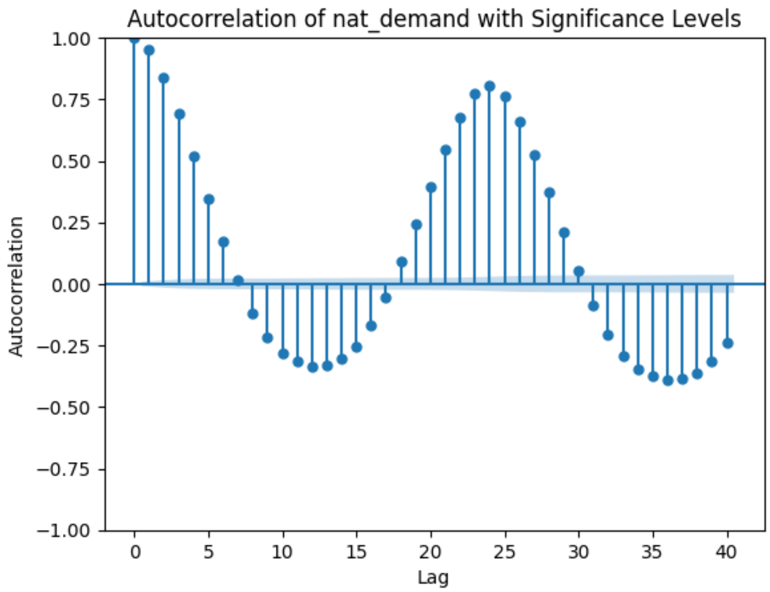
Autocorrelation Analysis

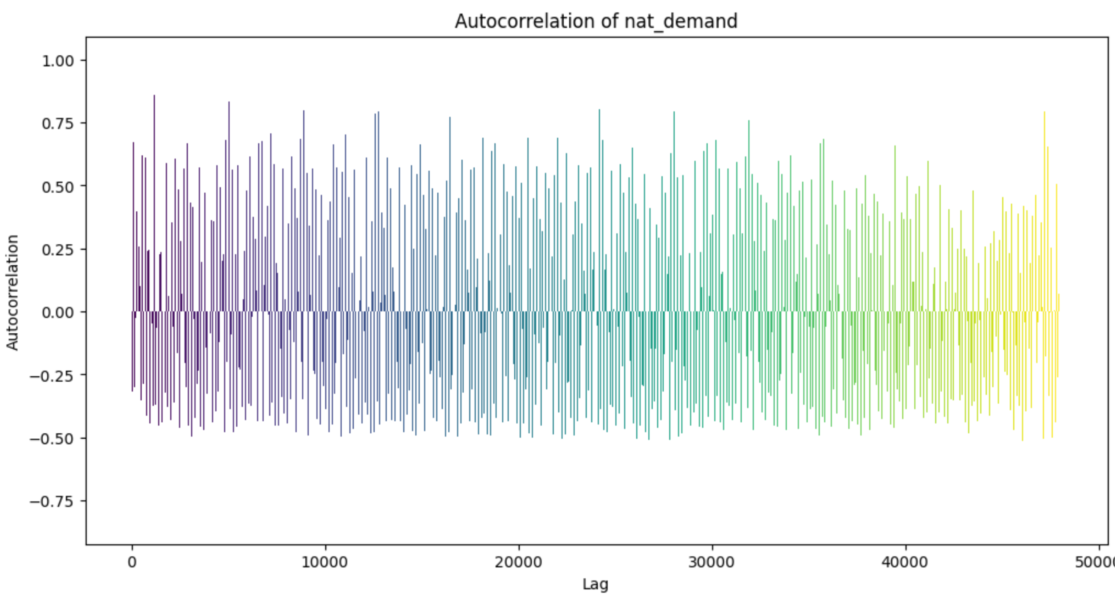
Autocorrelation is a statistical measure that assesses the relationship between a variable's current value and its past values. In time series analysis, autocorrelation helps identify patterns in the data. A high autocorrelation at a specific lag indicates a strong relationship between the variable's current value and its value at that lag. By analyzing autocorrelation at different lags, we can identify the presence of seasonal patterns and dependencies in the data.

To implement the AR model, we first analyzed the autocorrelation of the 'nat_demand' variable at different lags. Autocorrelation measures the relationship between a variable's current value and its past values, helping us identify patterns in the data. We used the autocorrelation function (ACF) plot to visualize these relationships and determine the number of lags to include in the AR model.

Next, we calculated the autocorrelation values for the 'nat_demand' variable at different lags and assessed their significance using confidence intervals. This step helped us understand the strength and significance of the relationships between past and future load values, guiding the selection of lag variables for the AR model.

Here is a sample of the autocorrelation graphs that shows lags Verses Autocorrelation:

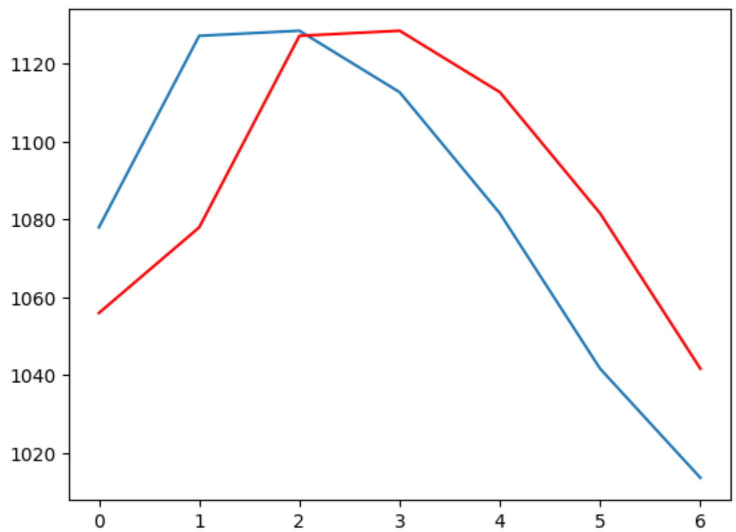




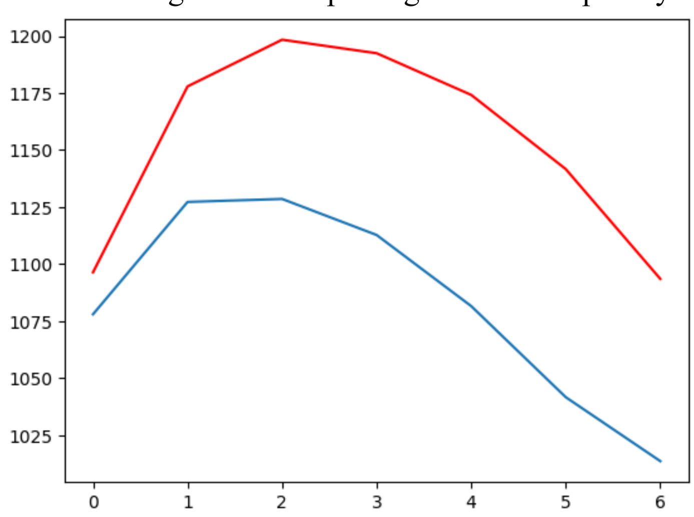
The graphs displays the autocorrelation of 'nat_demand' at various lags, indicating the strength and direction of the relationship between each observation and its lagged value. The blue bars represent autocorrelation coefficients, with the 95% confidence intervals shown as horizontal lines. A significant autocorrelation beyond these intervals suggests a pattern. For example, a spike at lag 24 indicates a strong correlation between current demand and demand 24 hours prior, hinting at a daily seasonal trend.

Persistence Model:
The persistence model, a basic time series forecasting technique, serves as a benchmark for evaluating more complex models. It simply uses the previous time step's value as the prediction for the current time step. While rudimentary, this model provides a foundational understanding of the dataset's characteristics and helps quantify the impact of its seasonal patterns on natural gas demand.

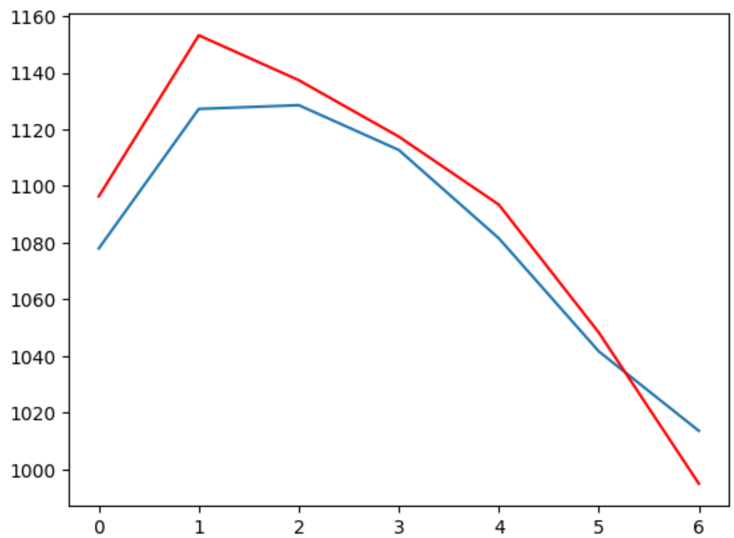
The model's performance, assessed using the mean squared error (MSE), indicates its limitations. With a test MSE of 931.311, the persistence model's predictive power is limited, primarily due to its inability to capture the underlying patterns and seasonality in the data. Its simplistic nature leads to a relatively high error rate, highlighting the need for more sophisticated modeling techniques.



Static Autoregressive Model:
The static autoregressive model represents an improvement over the persistence model by incorporating lagged values of the target variable as predictors. In this case, the model considers lagged values up to lag 29, providing a more nuanced understanding of how past observations influence future predictions.
The model's coefficients, representing the weights assigned to each lagged value, offer insights into the dataset's dynamics. Despite this improvement, the static nature of the model limits its adaptability to changing patterns over time. This limitation is reflected in its test root mean squared error (RMSE) of 74.761, indicating a moderate improvement over the persistence model but still falling short of capturing the full complexity of the data.



Updated Autoregressive Model:
The updated autoregressive model outperforms both the persistence and static autoregressive models by dynamically updating its coefficients at each time step based on the most recent data. This dynamic approach allows the model to adapt more effectively to changing patterns in the data, leading to more accurate predictions.
With a test RMSE of 15.339, the updated autoregressive model demonstrates significantly improved predictive power compared to the other models. By incorporating more recent information into its predictions, the model captures the evolving nature of seasonal patterns and variations in electricity consumption more effectively. This highlights the importance of dynamic modeling approaches in time series forecasting, especially when dealing with complex and evolving datasets.



In conclusion, the analysis of autoregressive models, including the persistence, static autoregressive, and updated autoregressive models, offers valuable insights into seasonal patterns in electricity consumption. The persistence model, though simple, provides a foundational understanding of seasonal trends but lacks the ability to capture the data's complexity, as seen in its high mean squared error (MSE). The static autoregressive model improves upon this by incorporating lagged values, providing a more nuanced view of how past observations influence future predictions. However, its static nature limits its adaptability to changing patterns over time.
The updated autoregressive model, with its dynamic approach to updating coefficients, outperforms both previous models. By incorporating recent information, it adapts more effectively to changing patterns, resulting in more accurate predictions. These models are crucial for energy providers, aiding in anticipating demand fluctuations, optimizing resource allocation, and ensuring a reliable electricity supply.

Conclusion:

In conclusion, the analysis using Facebook Prophet and autoregressive models has deepened our understanding of seasonal patterns in electricity consumption. Facebook Prophet has proven effective in capturing complex seasonalities and holiday effects, making it a valuable tool for short to medium-term planning. Complementing this, autoregressive models offer a detailed

perspective on how past observations influence future predictions, underscoring the significance of historical data in forecasting.
The Facebook Prophet analysis revealed consistent trends in electricity consumption, including seasonal, weekly, and daily patterns. The model demonstrated high accuracy, as evidenced by metrics like R-squared and RMSE, indicating its effectiveness in capturing these patterns and making reliable forecasts. Additionally, the trend analysis highlighted a steady increase in electricity consumption over time, emphasizing the need for proactive energy management.
On the other hand, the autoregressive models, including the persistence, static autoregressive, and updated autoregressive models, provided valuable insights into the underlying patterns in electricity consumption. While the persistence model acted as a baseline, the static and updated autoregressive models improved prediction accuracy by incorporating lagged values and adapting to changing patterns over time. These models serve as crucial tools for energy providers, aiding in anticipating demand fluctuations, optimizing resource allocation, and ensuring a stable electricity supply.

Moving forward, we can enhance seasonal pattern analysis in electricity consumption using Facebook Prophet and autoregressive models by refining the models, incorporating additional factors like weather data, exploring ensemble techniques, and investigating advanced modeling approaches. These steps can lead to more accurate predictions and better energy management strategies.

Conclusion

In this comprehensive analysis of electricity load forecasting, we explored multiple modeling techniques, including Facebook Prophet and autoregressive models, to understand seasonal patterns and variations in consumption. Each approach offered unique insights and advantages, contributing to a holistic understanding of electricity load forecasting.

Key Findings:

- Facebook Prophet: This model effectively captured seasonal patterns and holiday effects, demonstrating high accuracy in predicting electricity consumption. Its user-friendly design and ability to handle various seasonalities make it a valuable tool for short to medium-term planning.
- Autoregressive Models: The autoregressive models, including the persistence, static autoregressive, and updated autoregressive models, provided valuable insights into the underlying patterns in electricity consumption. While the persistence model served as a baseline, the static and updated autoregressive models improved prediction accuracy by incorporating lagged values and adapting to changing patterns over time.

Implementation in the Future:

- Ensemble Methods: Combining multiple models or simulation runs through ensemble techniques like bagging or boosting can further enhance forecasting accuracy and robustness.
- Uncertainty Analysis: Utilizing uncertainty analysis techniques, such as Monte Carlo simulations, can provide a more comprehensive understanding of potential outcomes and improve decision-making under uncertainty.
- Dynamic Modeling Approaches: Implementing dynamic modeling approaches, like the updated autoregressive model, can improve adaptability to changing patterns over time and lead to more accurate predictions.

Future Directions:

- Refining Models: Continuously refining and improving forecasting models to capture more nuanced patterns and variations in electricity consumption.
- Incorporating Additional Factors: Exploring the incorporation of additional factors such as weather data, economic indicators, and emerging trends to enhance forecasting accuracy.
- Advanced Modeling Techniques: Investigating advanced modeling techniques, such as deep learning or hybrid models, to further improve prediction accuracy and reliability.

In conclusion, by leveraging a combination of modeling techniques and uncertainty analysis, energy providers can gain valuable insights into seasonal patterns in electricity consumption, optimize resource allocation, and ensure a stable and reliable electricity supply. Continuous refinement and implementation of these techniques can lead to more accurate and efficient electricity load forecasting in the future.

References

Causal inference for the brave and true. 16 - Regression Discontinuity Design - Causal Inference for the Brave and True. (n.d.). https://matheusfacure.github.io/python-causality-handbook/16-Regression-Discontinuity-Design.html

Ot, A. (2023, November 14). *What is Data Simulation?: Benefits & modeling*. Datamation. https://www.datamation.com/big-data/data-simulation/#features

Rasmussen, J. T. (2022, January 9). *How to use an autoregressive (AR) model for time series analysis*. Medium. https://towardsdatascience.com/how-to-use-an-autoregressive-ar-model-for-time-series-analysis-bb12b7831024

Shafiuzzamansaikat. (2023, October 26). *FB prophet for load forecasting*. Kaggle. https://www.kaggle.com/code/shafiuzzamansaikat/fb-prophet-for-load-forecasting

Shap. (n.d.). *Shap/SHAP: A game theoretic approach to explain the output of any machine learning model.* GitHub. https://github.com/shap/shap

Using shap with pseven core for sensitivity analysis. pseven.io. (2021, April 27). https://www.pseven.io/blog/tech-tips/2021/using-shap-with-pseven-core-for-sensitivity-analysis.html