Smart Meter Project Combining IoT, AI and Machine Learning

EEM 602 – Term Project

1st semester of 2023-2024

Department of Electrical Engineering, University of Bahrain, Kingdom of Bahrain

Alawi Sharaf 202301262, Prof. Mohab Mangoud

Abstract:

This project aims to modernize the utilities industry by integrating advanced AIOT (Artificial Intelligence of Things) networks. Our primary goal is to enhance customer engagement and awareness regarding efficient electricity usage. The initiative is structured in two phases.

In the first phase, I focused on getting insights from smart meters. This enables a clearer understanding of consumer behavior and energy utilization patterns. By analyzing these insights, I aim to empower consumers with actionable information, fostering more efficient energy consumption and contributing to a greener footprint.

The second phase delves into enhancing the accuracy of electricity demand forecasting and implementing predictive maintenance for consumer appliances. Leveraging sophisticated machine learning models, including Linear Regression and Artificial Neural Networks, I analyzed smart meter data to predict future energy needs with high accuracy. This proactive approach not only ensures grid reliability but also significantly reduces maintenance costs by anticipating and addressing potential issues in consumer devices before they escalate.

Overall, this project represents a stride towards a more efficient, reliable, and customer-centric utility industry, marking a significant leap in smart energy management.

Introduction

This paper talks about how smart electricity meters and Artificial Intelligence (AI) are changing the way we use energy. First, I looked at how smart meters have developed over time. These meters have gone from being simple gadgets to really smart tools that can help manage energy better because of AI.

In this project, I didn’t build new meter hardware; I focused more on the ideas behind it. I used data from an earlier study to help with our work. This data helps us understand how people use electricity.

My main work was to use this data to see how two different computer methods – Linear Regression and Artificial Neural Networks (ANN) – can help us learn from this data. Both these methods have their own way of looking at data and comparing them shows us how each can be used to make energy use better.

At the end, I'll share what I learned from this project. I hope my findings can help make smart meters and energy management even better in the future.

In short, the paper is about exploring how AI can make smart meters more useful and how this can lead to better energy use.

Data Collection Infrastructure

In this section, I will explain the hardware setup that forms the backbone of our data collection process. Our approach builds upon the findings and methodologies detailed in a noteworthy study (1), which provides a robust framework for gathering smart meter data.

This previous research successfully deployed a two-part data collection system in a residential setting. The first part of this system involves smart meters capable of recording aggregated consumption data. These meters are designed to handle a maximum current of 50 Amperes and a voltage output of 1 Volt. Such specifications ensure the meters can accurately capture the overall energy usage in a household.

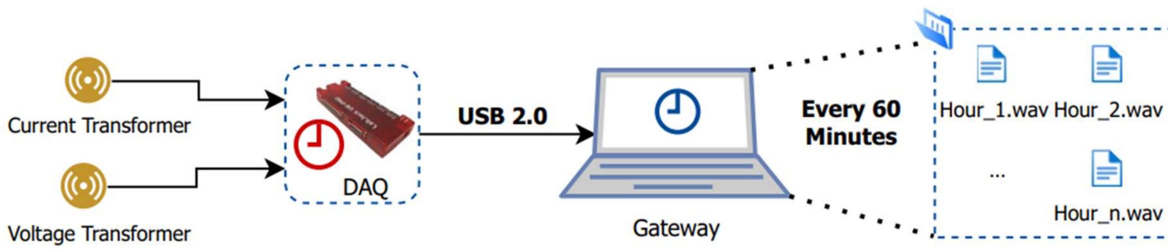


Figure 1: description: Main components of the data collection setup for the aggregated consumption (1)

The second part of the data collection involves monitoring the energy consumption of individual appliances. This is achieved using an innovative combination of Plugwise devices and Zigbee technology. Plugwise plugs, equipped with Zigbee wireless communication technology, are installed on various appliances. This setup allows for the accurate measurement of energy consumption by each appliance, transmitting this data for further analysis.

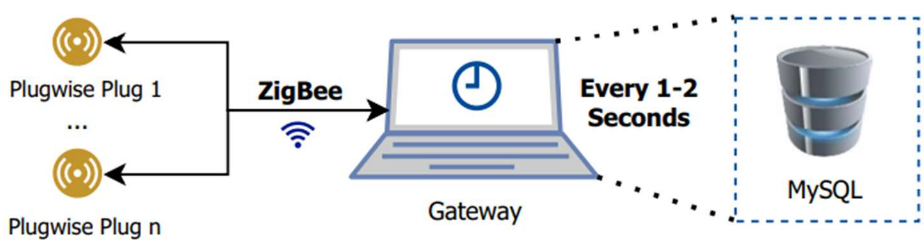


Figure 2: Main components of data collection setup for individual appliance consumption (1)

Figure 3: Some of the products that can be used to measure appliances usage

It's important to note that my study does not involve developing new hardware. Instead, I used the established infrastructure described above as my starting point. Photographs from the original paper illustrate this setup clearly.

The emphasis of our project is not on the hardware itself but on the analysis of the data collected through this setup. I aim to deep dive into this data, applying Machine Learning techniques – specifically Linear Regression and Artificial Neural Networks (ANN) – to draw meaningful insights. My comparison of these two methods forms the core of the research, offering a detailed understanding of energy consumption patterns and potential areas for efficiency improvements.

In summary, the hardware setup, as adopted from the referenced study, provides me with a comprehensive dataset, which is pivotal to my analysis and the subsequent conclusions I draw in our goal to enhance energy management and consumer engagement.

Data Exploration and Feature Engineering in MATLAB

A significant portion of this project was dedicated in preparing and analyzing the data in MATLAB, laying the foundation for our subsequent machine learning models.

1. Exploratory Data Analysis (EDA): my initial step involved a thorough exploration of the dataset. Utilizing MATLAB's robust analytical tools, I dived into understanding the nature of the smart meter data, which included variables like time stamps, active and reactive power, voltage, and current readings. Through a series of visualizations such as time series plots, histograms, and scatter plots, we gained valuable insights into the patterns and relationships inherent in the data. This process was not only about comprehending the data's structure but also about spotting any anomalies or outliers that could influence the model's performance.

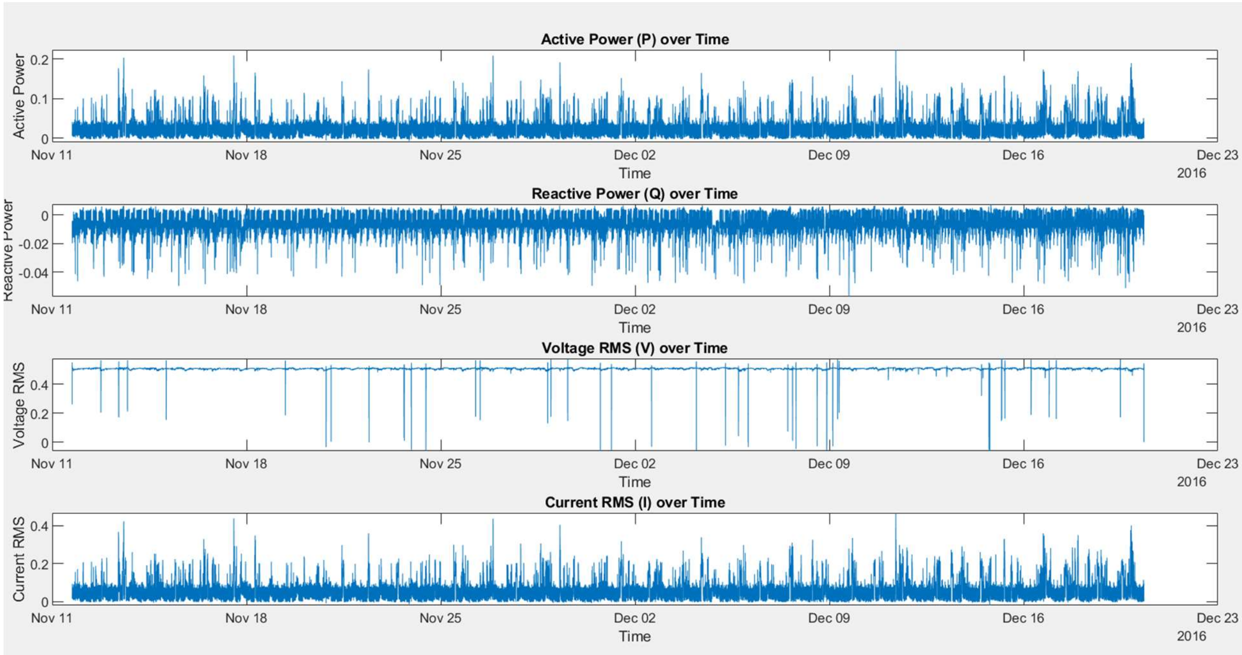


Figure 4: Time Serios Plot for all values (Active Power, Reactive Power, Voltage and Current)

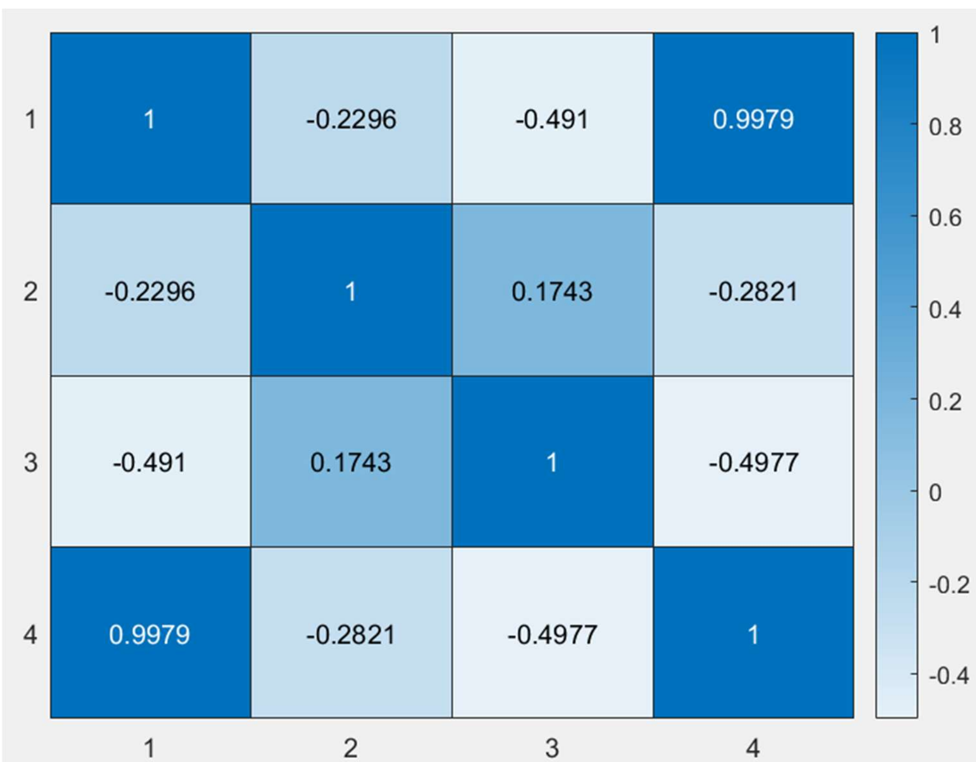


Figure 5: Correlation Analysis

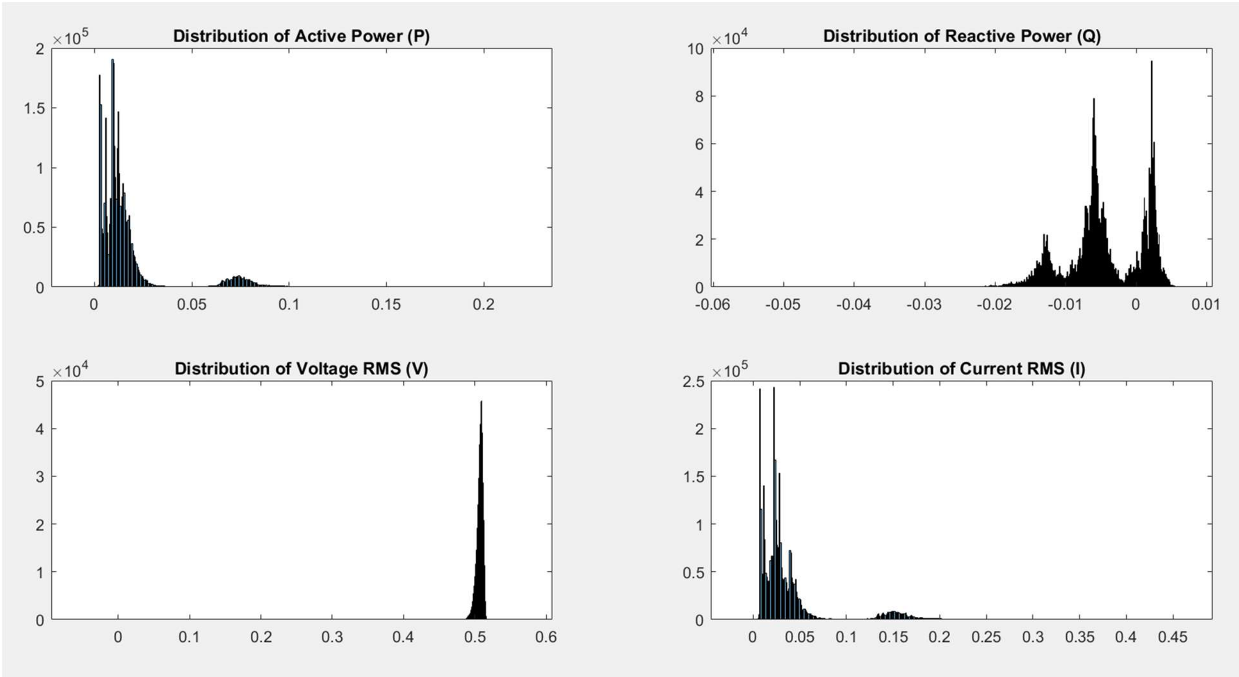


Figure 6: Histogram Distribution of each numerical Variable

2. Feature Selection and Engineering: Following EDA, I thought about the critical task of feature selection and engineering. This stage was key in transforming raw data into a format that can be used in machine learning. I identified and constructed features that would be most predictive of energy consumption patterns. This included transforming the time stamp data into more meaningful features like the hour of the day or the day of the week, which are more representative of consumer behavior. Additionally, I evaluated the correlation between different variables to streamline the features, ensuring that my models were fed with data that maximizes predictive power while minimizing redundancy.

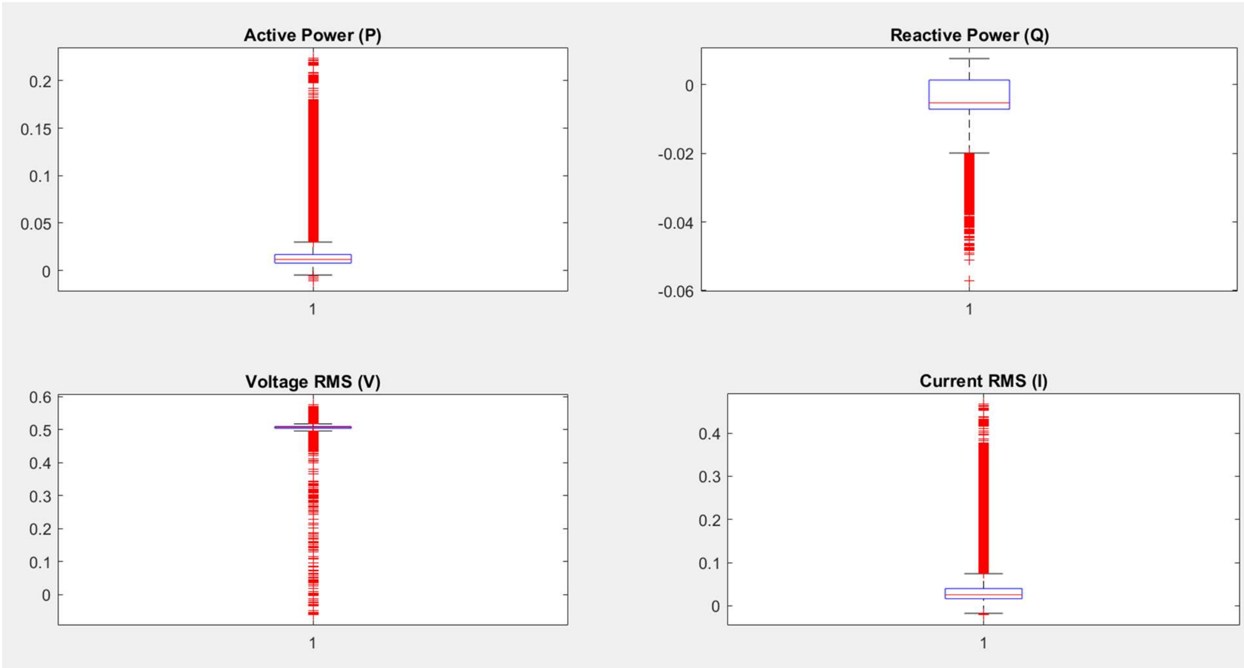


Figure 7: Box Plot Outliner in the data

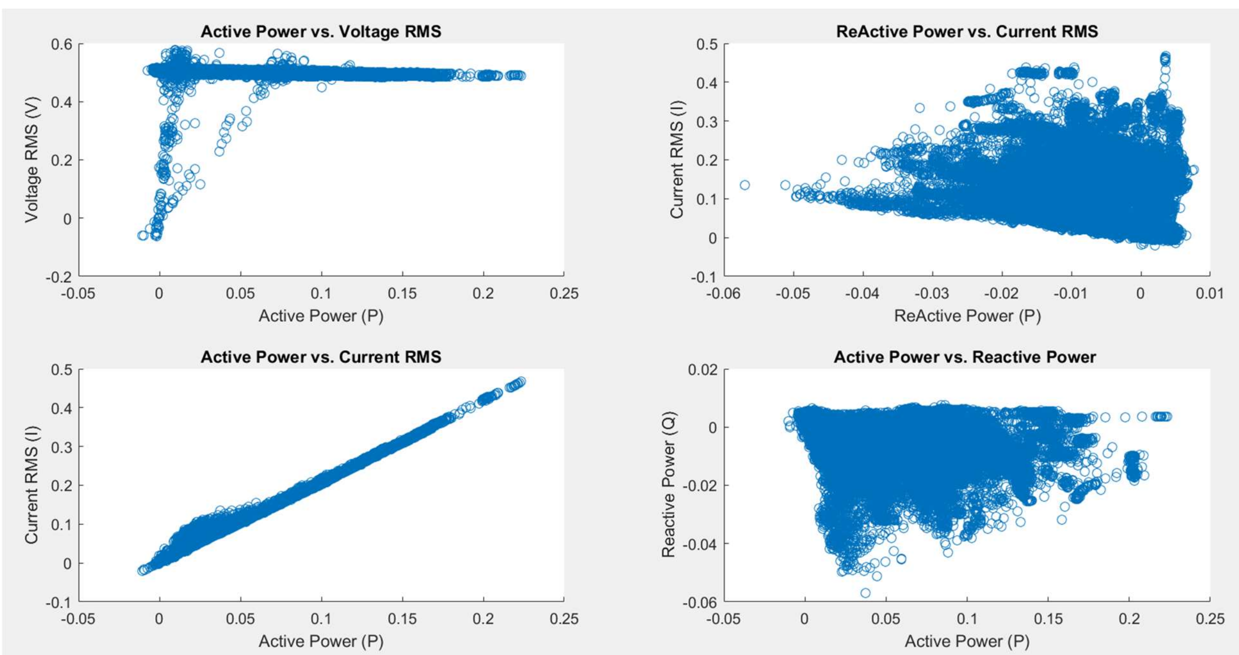


Figure 8: Scatter Plots for each data

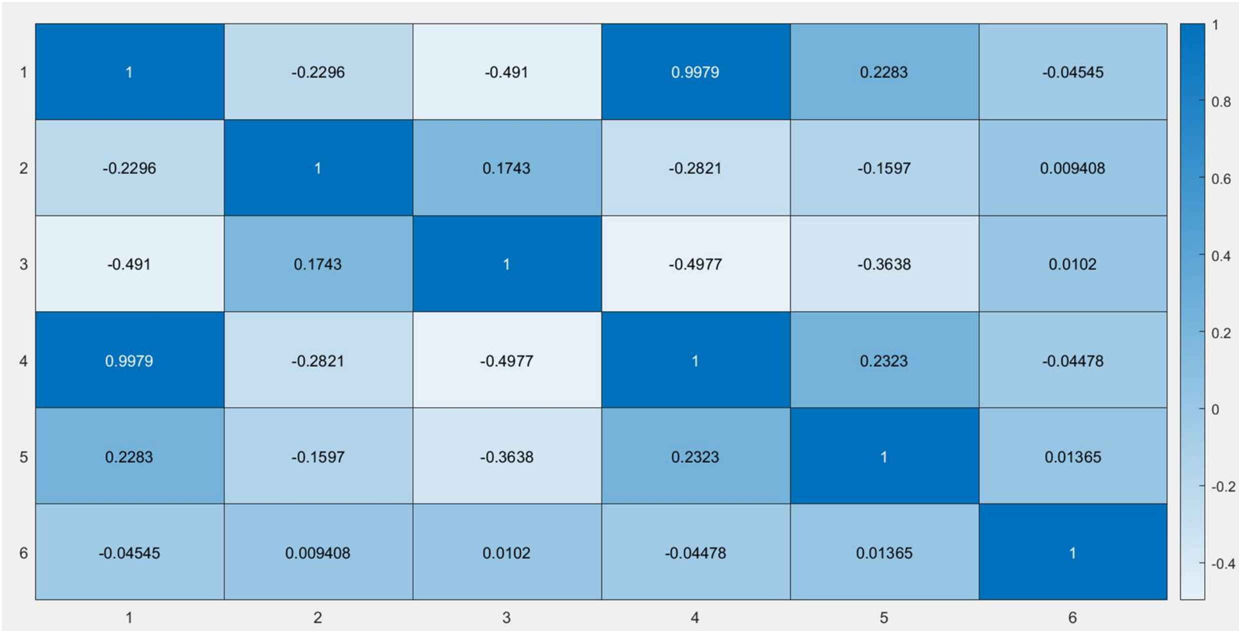


Figure 9: Revested Correlation analysis for all data

3. Model Selection and Training for Linear Regression: With the data suitably prepared and the features finely tuned, I first applied a Linear Regression model. The choice of Linear Regression was guided by its simplicity and interpretability, making it an ideal starting point for predicting energy consumption. I trained the model in MATLAB, leveraging its sophisticated statistical toolboxes. This phase was marked by careful calibration of the model to fit our data, followed by sharp training to ensure the model accurately captures the underlying relationships in the energy usage patterns.

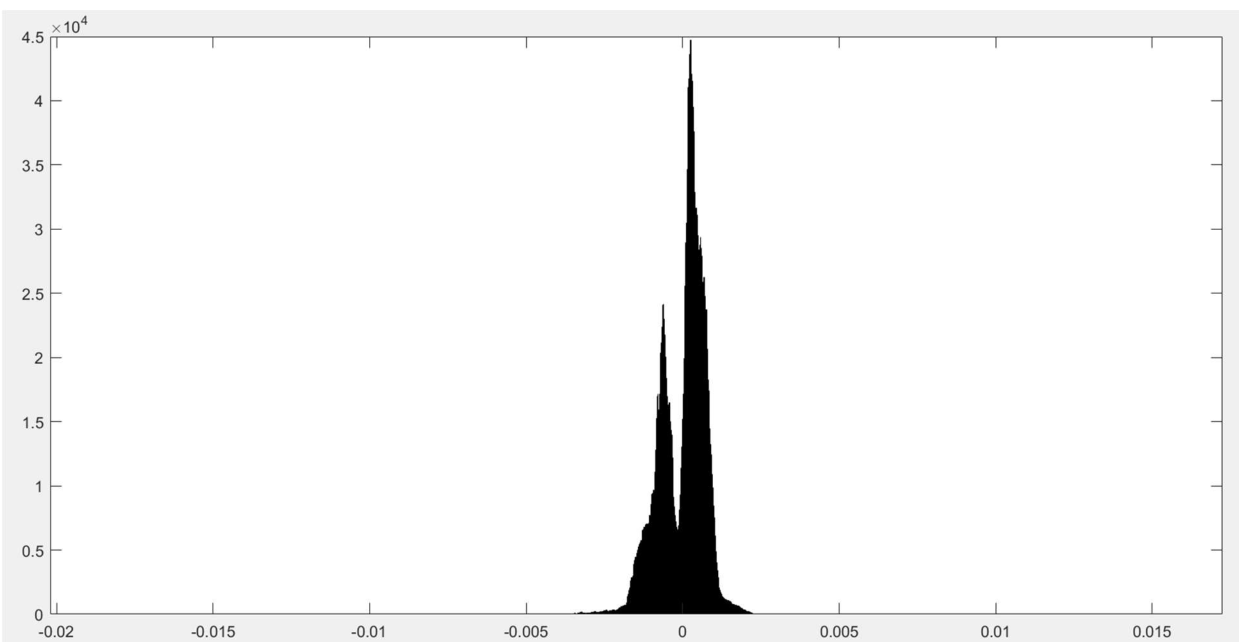


Figure 10: Linear regression result

```
Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5

Estimated Coefficients:
                   Estimate         SE          tStat       pValue
                   __________    __________    _______    __________

    (Intercept)     -0.013232    5.9273e-05    -223.24             0
    x1             1.0578e-05    7.5464e-08     140.17             0
    x2            -6.0275e-06    2.3656e-07    -25.479    3.4828e-143
    x3                0.21233    9.1632e-05     2317.2             0
    x4               0.024673    0.00011597     212.76             0
    x5                0.49558    1.3446e-05      36858             0


Number of observations: 2335694, Error degrees of freedom: 2335688
Root Mean Squared Error: 0.000743
R-squared: 0.999,  Adjusted R-Squared: 0.999
F-statistic vs. constant model: 3.69e+08, p-value = 0
Model coefficients:
   -0.0132
    0.0000
   -0.0000
    0.2123
    0.0247
    0.4956
```

Figure 11: Linear regression model result

The Linear Regression model I used shows a strong relationship between energy consumption and five key predictors. Notably:

- **Model Formula**: **y ~ 1 + x1 + x2 + x3 + x4 + x5**
- **Significant Predictors**: All predictors (x1 to x5) significantly influence energy consumption, with p-values near zero.
- **High Impact Variables**: x3, x4, and x5 show higher coefficients, indicating a strong impact on energy usage.
- **Model Accuracy**: The RMSE is 0.000743, suggesting the model's predictions are very close to the actual values.
- **Model Fit**: An R-squared value of 0.999 indicates the model excellently fits the data.

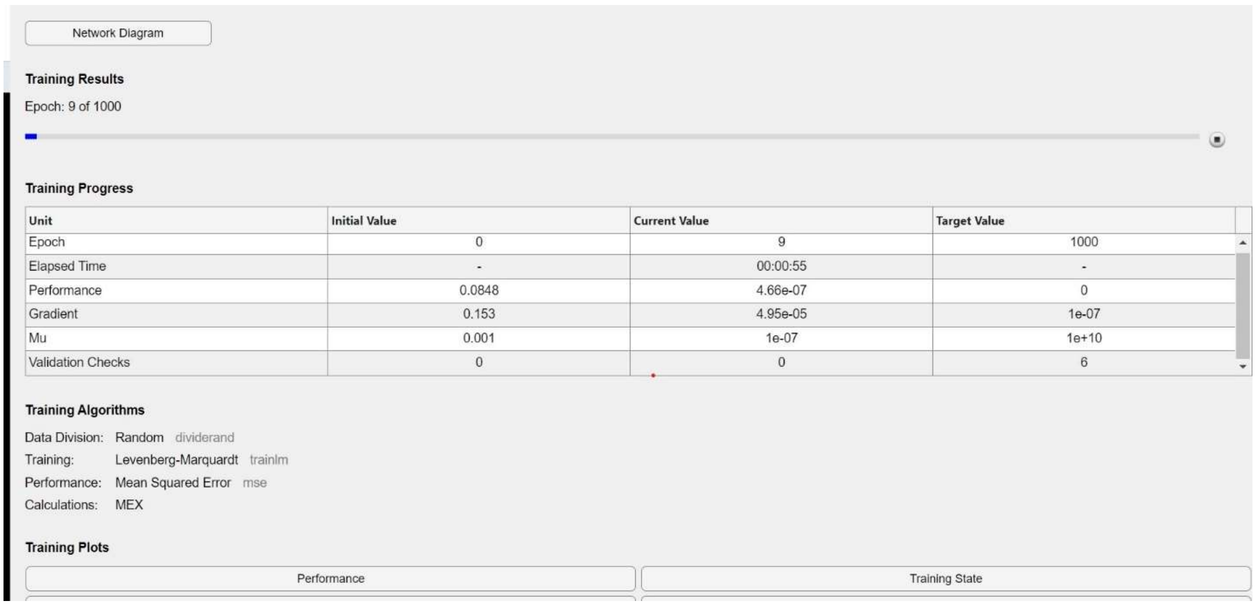
In short, the Linear Regression model effectively captures the dynamics of energy consumption, making it a valuable tool for understanding and predicting usage patterns.

4. Artificial Neural Network (ANN) Analysis

Following my analysis with Linear Regression, I shifted my focus to a more complex model, the Artificial Neural Network (ANN). This model was chosen for its ability to capture nonlinear relationships and intricate patterns in data, which might not be as apparent or accessible to linear models.

Data Preparation for ANN:

- The data preparation for the ANN was an extension of my earlier efforts in data exploration and feature engineering.
- I ensured the data was suitably formatted for ANN training, which involved transposing the data arrays to align with MATLAB's ANN input requirements.
- Special attention was given to normalize the input data, as ANNs are particularly sensitive to the scale of input values. This normalization process helps in improving the model's convergence and overall performance.



Network Diagram

Training Results

Epoch: 9 of 1000

Training Progress

Unit	Initial Value	Current Value	Target Value
Epoch	0	9	1000
Elapsed Time	-	00:00:55	-
Performance	0.0848	4.66e-07	0
Gradient	0.153	4.95e-05	1e-07
Mu	0.001	1e-07	1e+10
Validation Checks	0	0	6

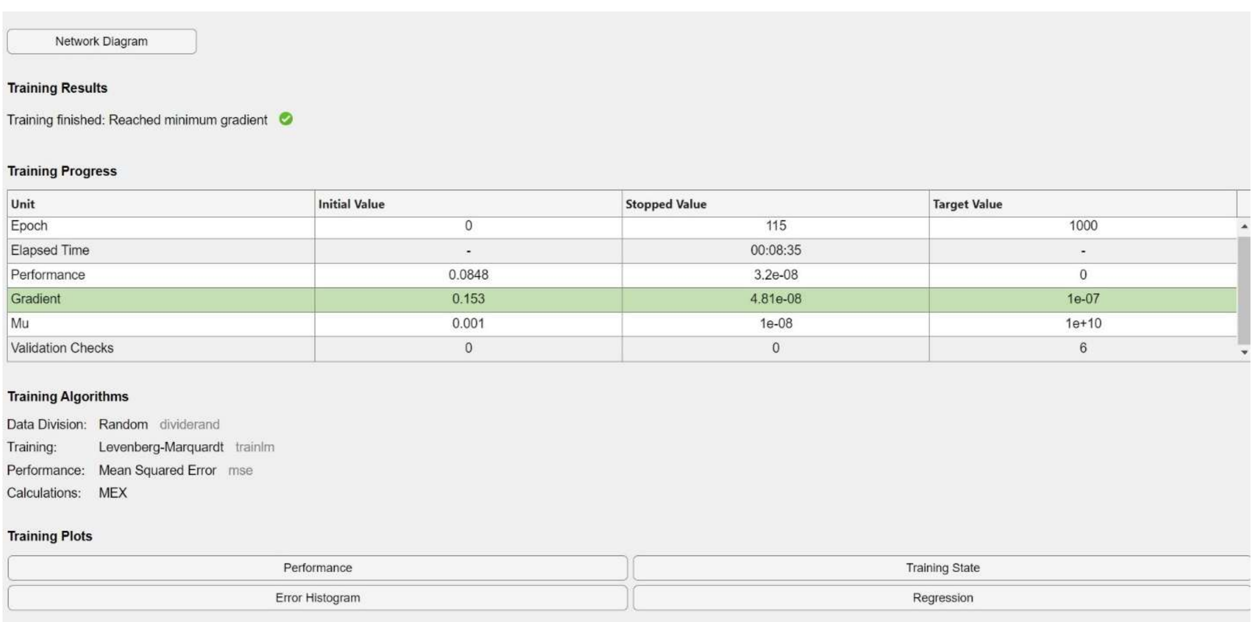
Training Algorithms

Data Division: Random dividerand
Training: Levenberg-Marquardt trainlm
Performance: Mean Squared Error mse
Calculations: MEX

Training Plots

Performance
Training State

Figure 12: ANN training data progress



Network Diagram

Training Results

Training finished: Reached minimum gradient

Training Progress

Unit	Initial Value	Stopped Value	Target Value
Epoch	0	115	1000
Elapsed Time	-	00:08:35	-
Performance	0.0848	3.2e-08	0
Gradient	0.153	4.81e-08	1e-07
Mu	0.001	1e-08	1e+10
Validation Checks	0	0	6

Training Algorithms

Data Division: Random dividerand
Training: Levenberg-Marquardt trainlm
Performance: Mean Squared Error mse
Calculations: MEX

Training Plots

Performance
Training State
Error Histogram
Regression

Figure 13: ANN ML Result

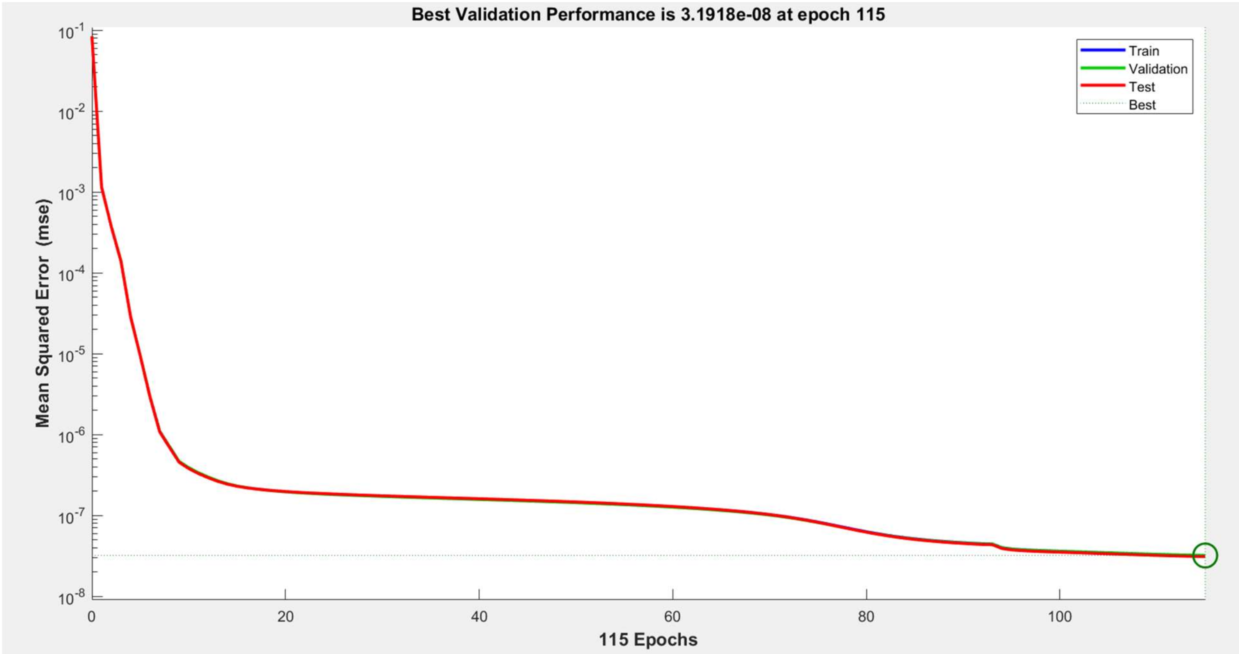


Figure 14: Best Validation Performance

ANN Model Training and Results:

- I constructed a feedforward neural network, carefully tuning parameters like the number of neurons and layers to strike a balance between model complexity and overfitting.
- The model was trained using the prepared dataset, and I monitored its performance throughout the training process, adjusting as needed.
- Upon training completion, our ANN model yielded impressive results. The model achieved a very high R-squared value of 0.99996, indicating an exceptional fit to the data.
- The RMSE for the ANN was calculated to be 0.00017937, signifying an excellent predictive accuracy and an improvement over the Linear Regression model.

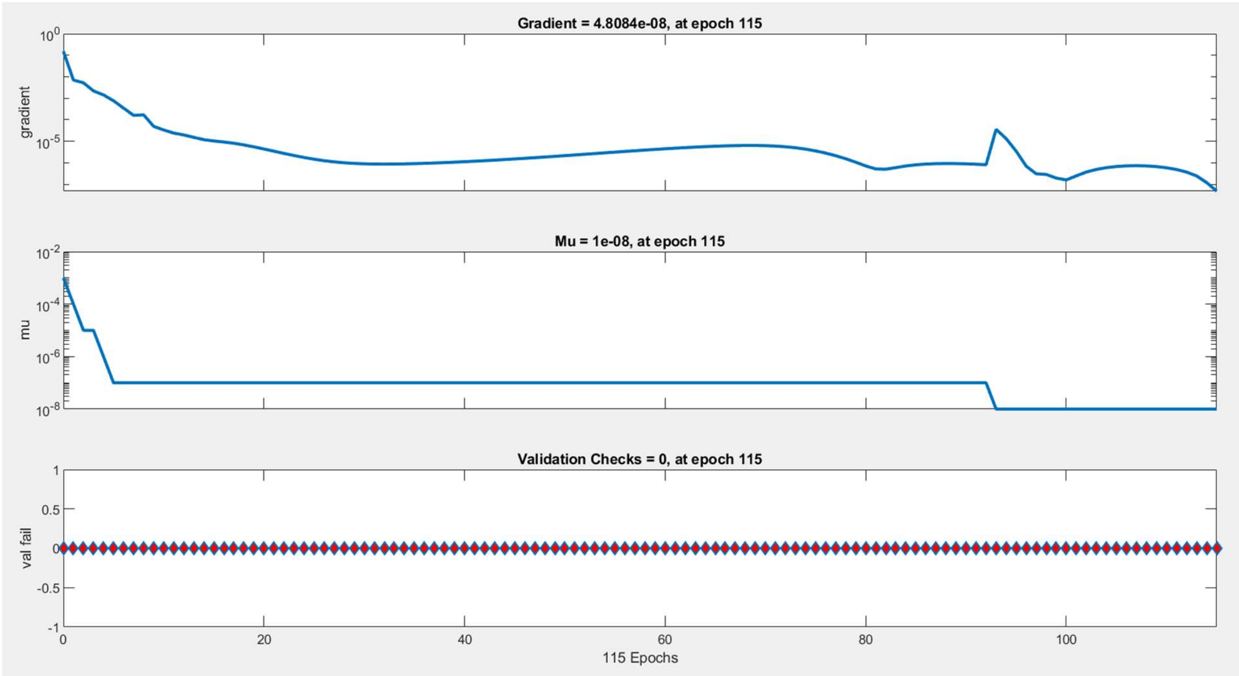


Figure 15: Neural Network Training State

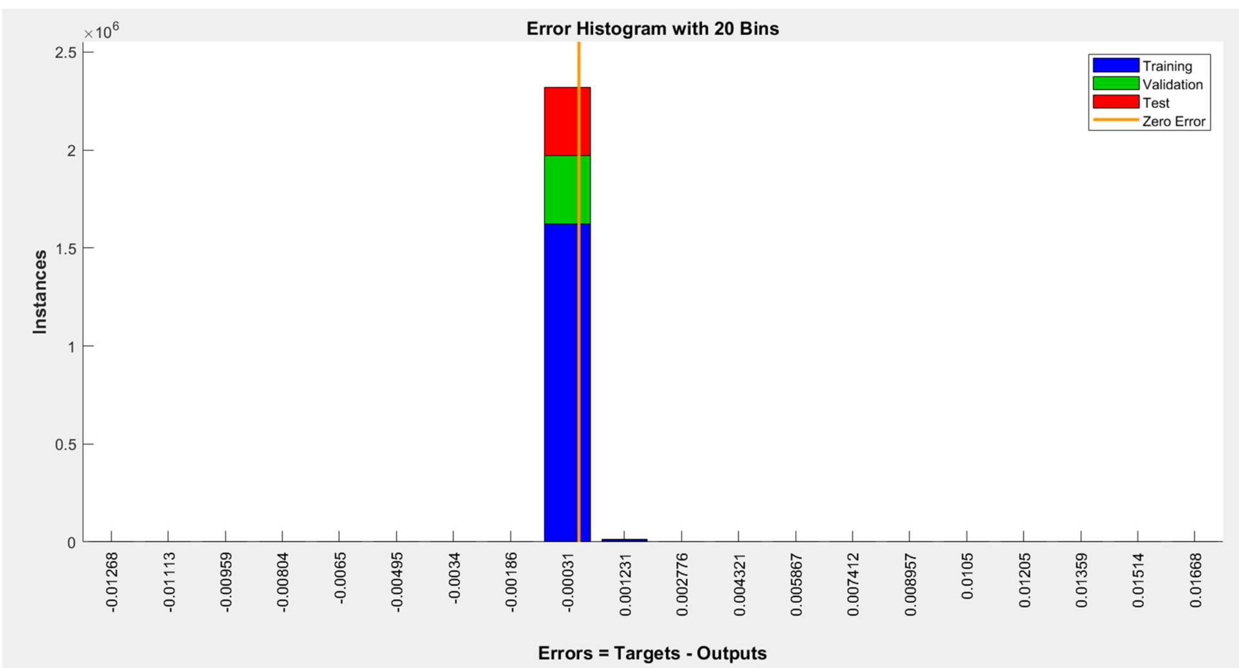


Figure 16: Neural Network Error Histogram

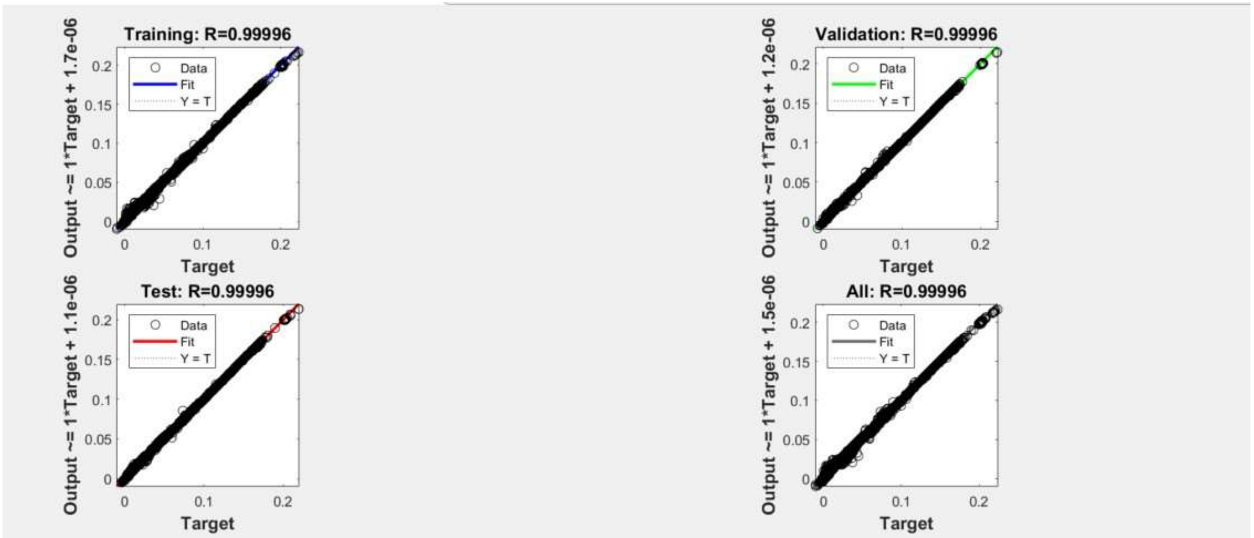


Figure 16: ANN Training Regression

The results show that the Artificial Neural Network (ANN) is really good at figuring out energy use patterns. It proves to be a strong tool for predicting how much energy will be used in the future. The ANN is great at understanding complicated data, which means it can give us a detailed look at how and when people use energy.

Methodologies and Calculations in Linear Regression and ANN

This section aims to provide a detailed insight into the methodologies and calculations used in my study, specifically focusing on the Linear Regression and Artificial Neural Network (ANN) models. A clear understanding of these concepts is crucial for interpreting the results and implications of our research.

Linear Regression Model:

Linear Regression is a statistical method that models the relationship between a dependent variable and one or more independent variables. In our case, we used it to predict energy consumption (dependent variable) based on various factors like time and energy readings (independent variables).

1. Model Formation:

- My Linear Regression model was formulated as **y ~ 1 + x1 + x2 + x3 + x4 + x5**, where **y** represents the energy consumption, and **x1** to **x5** are the predictor variables.

2. Coefficient Estimation:

- We used the Ordinary Least Squares (OLS) method to estimate the coefficients. This involves minimizing the sum of the squares of the differences between the observed and predicted values.

3. R-squared and Adjusted R-squared:

- R-squared (Coefficient of Determination): This metric indicates how well the regression predictions approximate the real data points. An R-squared of 0.999 means that 99.9% of the variation in the dependent variable can be explained by the independent variables in the model.

- Adjusted R-squared also considers the number of predictors in the model, adjusting for the number of variables.

4. RMSE (Root Mean Squared Error):

- RMSE is a standard way to measure the error of a model in predicting quantitative data. It represents the square root of the average squared differences between the predicted and actual values.

Artificial Neural Network (ANN) Model:

ANNs are computational models inspired by the human brain, capable of capturing complex patterns in data.

1. Network Architecture:

- I constructed a feedforward neural network, determining the optimal number of neurons and layers through iterative testing.

2. Training Process:

- The ANN was trained using backpropagation, where the network learns from the data by adjusting the weights to minimize prediction error.

3. Performance Metrics:

- Like the Linear Regression model, I calculated RMSE for the ANN to assess its predictive accuracy.
- R-squared value for the ANN was computed to determine how well the model predictions match the actual data. As with Linear Regression, a high R-squared value indicates a good fit.

Understanding these calculations and methodologies is vital for interpreting the outcomes of our models. The high R-squared values in both models suggest excellent fits to the data, with the ANN showing a slightly better performance in terms of lower RMSE. These models, with their respective strengths, offer significant insights into energy consumption patterns, highlighting the potential of machine learning in enhancing energy management and efficiency.

Reference to Prior Research in Smart Meter Data Analytics

In the world of AMI (Advanced Metering Infrastructure) (2) smart meter data analytics, significant studies have been made, as evidenced by notable research in the field. One such study, focusing on the analysis of 5-minute smart meter data sets from 100 anonymized commercial buildings, stands out for its comprehensive approach to electricity consumption patterns. This research highlights the immense potential of smart meters in generating vast volumes of data, presenting opportunities for both utilities and consumers. Utilities can enhance customer service, reduce costs, and improve energy efficiency, while consumers can benefit from reduced bills and energy savings.

This study emphasizes the complex nature of smart meter data analytics, encompassing data ingestion, pre-processing, analysis, and visualization. It dives into the exploration of time series of electricity consumption and compares various forecast models using a similar day approach. By conducting analytics on sub-industry levels and employing predictive analysis, the study enhances forecast models, particularly highlighting the effectiveness of ARIMA and exponential smoothing techniques, which resulted in lower prediction errors.

While acknowledging the insights and methodologies of this research, my study opts for a different analytical path. I focus on exploring machine learning techniques, particularly Linear Regression and Artificial Neural Networks (ANN), to analyze smart meter data. This choice is driven by my aim to compare the efficacy of these two distinct methodologies in understanding and predicting energy consumption patterns, thereby contributing a new perspective to the field of smart meter data analytics.

Conclusion

This project is right at the cutting edge of the fast improvements in technology for utility companies, especially for smart meters. It all started by looking at how much better smart meters have gotten, which really helps us analyze data and connect better with customers.

The main part of my project was to compare two different ways of using computers to predict things: Linear Regression and Artificial Neural Networks (ANN). The Linear Regression method was easy to understand and gave us really accurate results, with a score of 0.999 and an error rate of just 0.000743. The ANN was even better at dealing with complicated data, scoring a bit higher at 0.99996 and having an even lower error rate of 0.00017937. These results show that both methods are great at predicting how much energy people will use and help us understand customer habits better.

What we've found is really important. By using these prediction methods, we can get a clearer picture of how people use energy. This helps companies manage energy better and keeps the power grid reliable. It also helps customers use less energy, which saves them money and is better for the planet.

But this project is just the start. There's a lot more we can do with machine learning in the energy industry. We've just begun exploring, and there are many other techniques to try out. As we keep moving forward in our AI work, these first discoveries will help us do even more research and development. In future projects, we might look at other models, dig deeper into the data, or even use real-time predictions to make the energy industry smarter, more efficient, and focused on what customers need.

References:

[1] L. Pereira, D. Costa, and M. Ribeiro,(2022) "A Residential Labeled Dataset for Smart Meter Data Analytics, Scientific Data

[2] Mohammad Harun Rashid. (2018). AMI Smart Meter Big Data Analytics for Time Series of Electricity Consumption. IEEE International Conference

[3] Kardi, M; Al Saif, Tekinerdogan, B; Catala, JoaoP.S.(2021). Anomaly Detection in Electricity Consumption Data using Deep Learning. IEEE International Conference

[4] Gubbi, J., et al. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. ScienceDirect.