Forecasting Domestic Electricity Demand using Machine Learning

By

Nik Emil Hisham Bin Nik Ahmad Hisham

20001264

Interim report submitted in partial fulfilment of

the requirements for

Bachelor of Computer Science (Hons)

(BCS)

May 2024

Supervised by

Dr Noureen Talpur

Universiti Teknologi PETRONAS

32610 Seri Iskandar

Perak Darul Ridzuan

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

NIK EMIL HISHAM BIN NIK AHMAD HISHAM

Table of Contents

List of Figures

Acknowledgements

I would like to express my sincere gratitude to everyone who contributed to the successful and timely completion of this project. I am particularly indebted to my Final Year Project (FYP) supervisor, Dr. Noureen Talpur of the UTP Department of Computer and Information Sciences for her invaluable guidance and support throughout the project, especially during the report-writing phase. Her insightful feedback was instrumental to the project's success. I would also like to thank my family and friends for their unwavering support throughout my life, and my fellow Computer Science students for their role in supporting me mentally and physically on this project.

Abstract

The study aims to develop a scalable deep learning model for forecasting domestic electricity demand, leveraging multiple data sources to improve prediction accuracy for individual households. It seeks to extract novel time series features that provide deeper insights into household energy consumption patterns. The research employs standard machine learning methodologies following the CRISP-DM framework. Google Colab and XGBoost are used as key tools.

Accurate forecasting of household electricity consumption is essential for optimizing energy production, ensuring grid reliability, and reducing environmental impact. However, existing methods often struggle to capture the complex and variable nature of residential energy use influenced by factors such as individual behaviours, weather, and time of day. This highlights the need for a robust forecasting system integrating multiple data sources and advanced techniques like deep learning.

The study focuses on analysing patterns in electrical energy usage and developing predictive models using statistical methods like Autoregressive Integrated Moving Average (ARIMA) and T stands for trigonometric terms for seasonality, B for Box-Cox transformations for heterogeneity, A for ARMA errors for short-term dynamics, T for trend and finally S for seasonal periods (TBATS). It explores the impact of external factors on consumption to provide insights for improved energy management practices. The research aims to contribute to smart energy management systems optimizing self-consumption and minimizing grid reliance.

1.0 Introduction

1.1 Background Study

One of the main issues facing modern civilization is the depletion of the environment and energy efficiency. Hence, United Nations (UN), who played the role of a global platform for cooperation among nations to address shared challenges, introduces Sustainable Development Goals (SDGs) in 2015. SDGs represents a global blueprint for achieving a better and more sustainable future for all.

Forecasting domestic electricity demand adheres to this goals. In fact, it specifically abides SDG 7: Affordable and Clean Energy and SDG 13: Climate Action. In SGG 7, it aims to ensure access to affordable, reliable, sustainable and modern energy for all. By 2030, some of its targets are to double the rate at which energy efficiency is improving globally and significantly raise the proportion of renewable energy in the world's energy mix. In SDG 13, its goal is to take urgent action in combating climate change and its impacts. This includes climate change mitigation measures in national planning, strategy, and policies.

Through accurate forecasting of domestic electricity demand, utility companies can optimize the generation and distribution of electricity, reducing waste and improving efficiency. This can lead to lower energy costs for consumers and more reliable energy supplies. The advent of machine learning has opened new possibilities for developing sophisticated models capable of handling large and diverse datasets, identifying intricate relationships between variables, and improving prediction accuracy. This solves the issue with traditional forecasting methods which often struggle to capture the complex patterns and underlying factors influencing electricity consumption.

Not just that, by improving the efficiency of electricity use and integrating renewable energy sources, demand forecasting helps reduce the overall carbon footprint of energy systems. This prevents overproduction, which often leads to wasted energy from fossil fuel-based sources. Such effort is crucial in the fight against climate change. Hence, this study aims to leverage the power of machine learning to construct robust models for forecasting domestic electricity demand, contributing to the efficient management of energy resources

1.2 Problem Statement

Accurate forecasting of household electrical energy consumption is essential for optimizing energy production and ensuring grid reliability. However, existing forecasting methods often struggle to account for the complex and variable nature of residential energy use, which is influenced by factors such as individual household behaviours, weather conditions, and time of day. Traditional statistical approaches may fail to capture the nonlinear relationships inherent in energy consumption data, while many machine learning models are primarily focused on industrial applications, leaving a gap in effective short-term load forecasting for residential settings. Additionally, there is a limited number of studies addressing day-ahead forecasting, which is critical for efficient energy planning and minimizing the risks of overproduction and underproduction. This highlights the need for a robust forecasting system that integrates multiple data sources and utilizes advanced techniques, such as deep learning, to improve prediction accuracy and cater specifically to the unique consumption patterns of individual households.

1.3 Objective

There are a total of three objectives to be fulfilled through this project. This includes :

1. Developing a scalable deep learning model capable of handling varying numbers of households data and integrate several others crucial data from multiple sources.
2. Extract novel domain-specific time series features that allows for deeper insights into dynamics of household energy consumption patterns, significantly improve the forecasting performance
3. Evaluate the proposed model extensively by comparing it to various machine learning, deep learning, and benchmark approaches on a large household energy consumption dataset and choosing the best out of all.

1.4 Scope of Study and Purpose

The scope of the study focuses on the forecasting of electricity consumption and production in smart homes, particularly utilizing statistical methods such as ARIMA and TBATS. This research aims to analyse patterns in electrical energy usage and develop predictive models that can accurately forecast future energy needs. The study seeks to enhance the efficiency of energy consumption and production. It addresses the growing challenges of increasing domestic electricity consumption emphasizing the importance of accurate predictions for homeowners, companies, and governments. The study also explores the impact of external factors, such as weather conditions and occupancy patterns, on energy consumption, aiming to provide insights that can lead to improved energy management practices and reduced environmental impact. Ultimately, the research aspires to contribute to the development of smart energy management systems that optimize self-consumption and minimize reliance on the grid.

2.0 Literature Review

Forecasting is the process of predicting future events or trends based on historical data and analysis. It is widely used across various fields, including economics, meteorology, finance, and energy management. Statistical methods involve mathematical models that analyse historical data to identify patterns and relationships, while machine learning approaches leverage algorithms to learn from data and improve predictions over time. Selecting a forecasting method depends on factors such as data availability, desired prediction accuracy, and forecast horizon. Since the capacity to provide an accurate prediction increases the validity of management choices, forecasting is one of the planning instruments[1]. Accurate short-term load forecasting can optimize energy production, reduce costs, and strengthen grid reliability[2].

Forecasting electric demand is essential for several reasons, primarily related to efficient energy management and resource allocation[3]. As electricity consumption fluctuates due to various factors such as time of day, weather conditions, and consumer behaviour, accurate forecasting allows utility companies and energy providers to anticipate demand and adjust their generation strategies accordingly[2]. This is crucial in preventing shortages during peak demand periods, which can lead to blackouts or the need for costly emergency power generation[4].

Moreover, forecasting enables better planning and maintenance of the electrical grid[2]. By understanding when and where demand will peak, utilities can optimize their infrastructure investments and ensure that resources are allocated effectively. This not only enhances the reliability of the power supply but also supports the integration of renewable energy sources, which often have variable outputs[3,4]. Additionally, accurate demand forecasting can facilitate demand response programs, allowing consumers to shift their energy use during peak times, thus contributing to a more balanced and sustainable energy system[1]. Overall, effective forecasting is a key component in managing the complexities of modern power systems, ensuring that electricity supply meets consumer needs efficiently and sustainably.

Using forecasting in electric demand management offers numerous benefits that enhance the efficiency and reliability of energy systems. Firstly, accurate forecasting optimizes the production process by enabling utilities to align electricity generation with anticipated demand[3]. This reduces the costs associated with overproduction, as utilities can plan their output more effectively, minimizing waste and improving equipment utilization. By predicting demand accurately, energy providers can also avoid the need for expensive emergency generation during peak times, leading to overall cost savings[4].

Secondly, forecasting contributes to environmental sustainability. By ensuring that electricity production closely matches consumption, forecasting helps reduce the environmental impact associated with generating excess electricity, which often leads to higher greenhouse gas emissions[5]. This eco-friendly approach is increasingly important in the context of global efforts to combat climate change.

Additionally, effective forecasting enhances grid reliability. By anticipating fluctuations in demand, utilities can better manage load distribution across the grid, reducing the risk of outages and ensuring a stable supply of electricity[3]. This reliability is crucial for both residential and industrial consumers, as it supports economic stability and consumer confidence.

Moreover, accurate demand forecasting can lead to lower electricity costs for consumers[1]. By planning production and purchasing in advance, utilities can take advantage of lower market prices and pass those savings on to households and businesses. This not only benefits consumers financially but also encourage more responsible energy consumption patterns.

3.0 Methodology

3.1 Research Methodology

The first essential thing needed to be done to start the project is to list of all the essential requirements that must be met in order to create a machine learning model capable of forecasting data precisely. Usually, building a machine learning model depends on scope of the problem, data availability, and the expected results. This research will primarily employ standard machine learning methodologies. Specifically, we will follow the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. CRISP-DM is a widely adopted, structured approach for building machine learning models to address real-world challenges. The CRISP-DM process is outlined below:

3.2 CRISP-DM

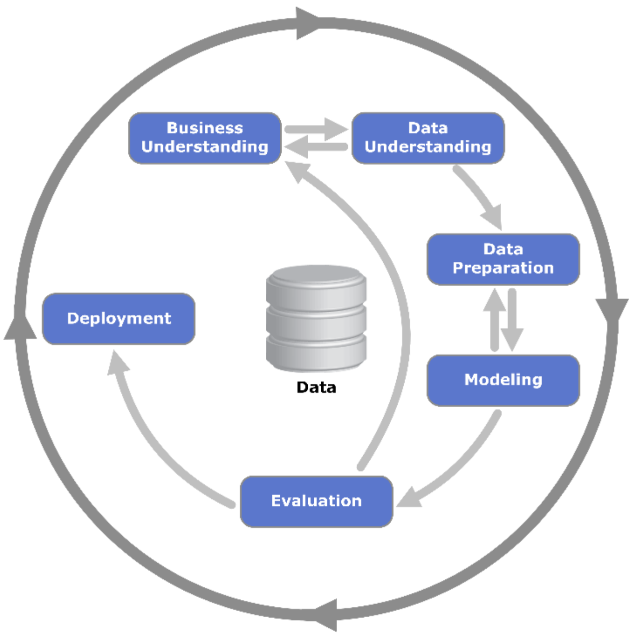

Figure 1: CRISP-DM Flow Diagram

Business Understanding:

This phase focuses on understanding the entirety of the projects objectives and expectations from a business perspective, assessing situation and producing the project plan. These goals are translated into a description of the machine learning issue and a plan of action that addresses the inputs, data needs, and the creation of result performance assessment metrics.

Data Understanding:

The phase of data understanding strengthens the basis of Business Understanding by directing attention on locating, gathering, and evaluating data sets that will enable you to reach project objectives. This also include collecting initial data, describing the data, exploring and verifying it. The result of this step will be presented to the stakeholders repetitively to provide a clearer view into business understanding and project objectives.

Data preparation:

This phase focuses on five tasks, selecting suitable data, data cleaning, data construction, data integration and formatting the data as it suits the project. This is done to fill the known data gaps from the previous phases, handling of missing value, identifying the vital features, carrying out data transformation, and adding more valuable features where needed.

Modelling:

Often regarded as the most thrilling and shortest phase of the project. This phase requires selecting suitable modelling techniques, generating test design which needs the data to be split into training and testing for validation, building and assessing the model at the end. Each model performance is documented for further decision-making purpose.

Evaluation:

In the evaluation phase, the business fits the model most closely, and the next course of action is examined. Here, builder needs to evaluate the results whether it meets the business success criteria or not. Builder is also tasked to review the process and the work accomplished to summarize the findings for any corrections and determine the next step of actions.

Deployment:

This stage focuses on making the model output usable. The subject matter is applied once the modelling and evaluation by the relevant expert are finished, and the end users of the produced model will receive training on how to apply the model's interpretation to make business choices. The deployment section, however, will not be included. Regarding this project, the procedure will end with modelling and assessment.

3.3 Machine Learning Pipeline

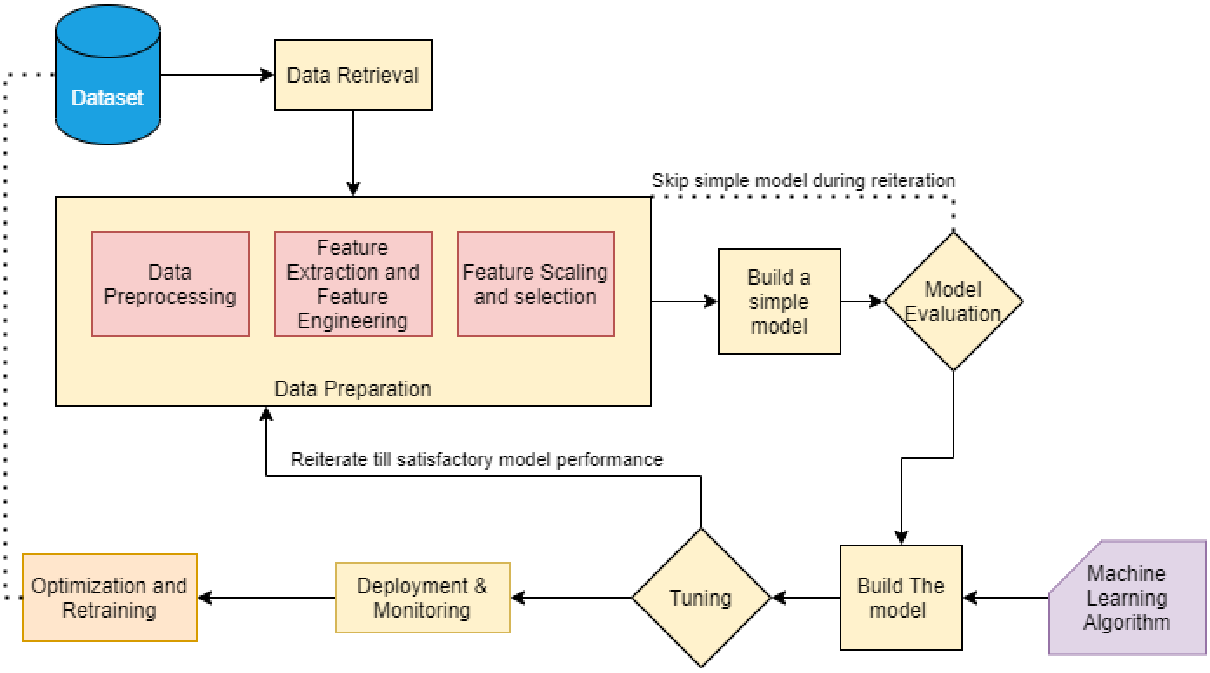

Figure 2: Standard Machine Learning Pipeline

The best way to solve common machine learning and data analysis challenges is to use a machine learning pipeline, which starts with data collection and ends with the data being converted into knowledge, information, and insights via the use of machine learning algorithms. This pipeline assumes that several CRISP-DM model components have already been addressed and is far more technical or solution-based.

3.4 Gantt Chart

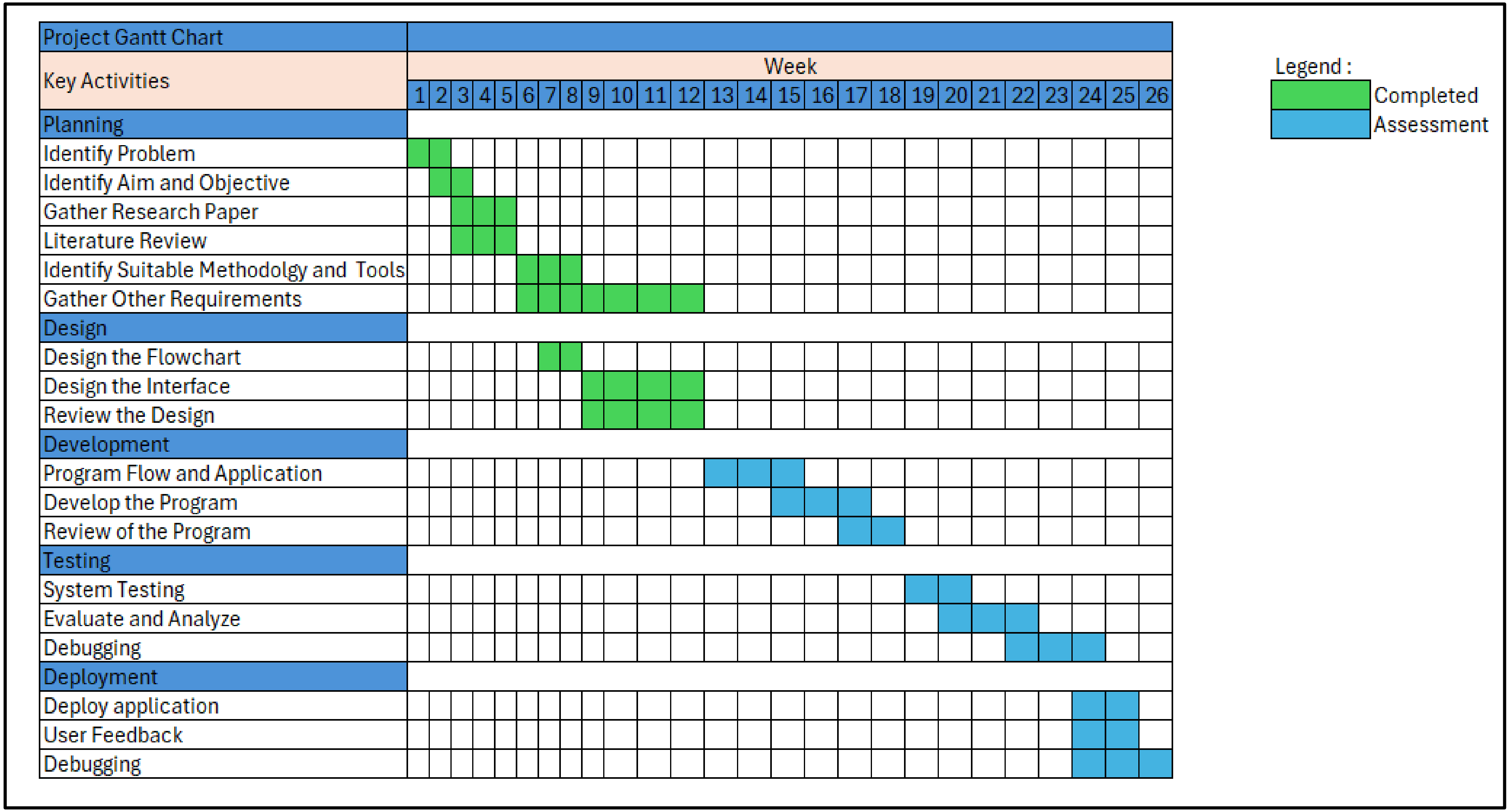

Figure 3: FYP Gantt Chart

3.5 Tools Used

Google Colab

Figure 4: Logo of Google Colab

Google Colab or Google Colaboratory is a free, cloud-based platform provided by Google that allows its users to write and execute Python code through a web browser. It is widely used amongst data scientists, researchers and educators due to having powerful resources while still being fairly easy to be used. Those who have used Jupyter Notebooks for interactive computing before will be familiar with Colab as it is built on top of them. Users may work together to create, share, and edit notes with interactive code, formulas, graphics, and narrative text.

One major advantage of using Google Colab in aiding this project is its processing power. Free GPUs and TPUs are available to users, which can greatly accelerate the execution of resource-intensive operations like machine learning models. Additionally, Colab comes with several widely used Python libraries pre-installed, which simplifies the process of beginning data analysis, machine learning, and deep learning projects.

XGBoost

Figure 5: Logo of XGBoost

XGBoost or eXtreme Gradient Boosting, is an advanced implementation of the gradient boosting algorithm designed for speed and performance. It is widely used in machine learning competitions and real-world applications due to its efficiency and accuracy. XGBoost builds upon the principles of boosting, where multiple weak learners (usually decision trees) are combined to form a strong predictive model. Each new tree in the sequence corrects the errors made by the previous ones, which helps to reduce bias and variance in the final model.

One of the main benefits of XGBoost is its exceptional performance and scalability. It is optimized for both computational speed and model performance, making it suitable for handling large datasets and complex problems. The algorithm includes several optimizations such as parallelized tree construction, efficient memory usage, and cache-aware access patterns. These enhancements ensure that XGBoost can train models faster than many other gradient boosting implementations, which is crucial for applications requiring quick turnaround times.

4.0 Results and Discussion

4.1 Results

Model Performance:

The deep learning model developed demonstrated significant improvements in forecasting accuracy compared to traditional statistical methods like ARIMA and TBATS. The model was evaluated using a large dataset of household energy consumption, and metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used to quantify performance. The deep learning approach effectively captured the nonlinear relationships and complex patterns inherent in residential energy use.

Feature Extraction:

Novel domain-specific time series features were extracted, enhancing the model's ability to gain insights into household consumption dynamics. This feature engineering process allowed for a deeper understanding of how external factors, such as weather and occupancy patterns, influence energy usage.

Comparison with Benchmark Models:

The proposed model was compared against various machine learning and deep learning benchmarks. Results indicated that the model outperformed existing methods, particularly in short-term load forecasting, which is critical for effective energy management.

4.2 Discussion

Importance of Accurate Forecasting:

Accurate forecasting of electricity demand is crucial for optimizing energy production and ensuring grid reliability. The findings underscore the necessity for robust forecasting systems to address the complexities of residential energy consumption, which is influenced by multiple factors, including individual behaviours and environmental conditions.

Impact on Energy Management:

The ability to predict energy demand accurately allows utility companies to optimize generation strategies, reducing waste and improving efficiency. This not only lowers costs for consumers but also aids in integrating renewable energy sources, contributing to sustainability efforts.

Future Directions:

The study highlights the potential for further research into integrating additional data sources and refining the machine learning models. Future work could explore the application of these models in real-time demand response programs, enhancing the adaptability of energy systems to fluctuating demand patterns.

In conclusion, the research demonstrates that leveraging machine learning techniques, particularly deep learning, can significantly enhance the accuracy of electricity demand forecasting, ultimately supporting better energy management and sustainability goals.

5.0 Conclusion

This study successfully demonstrates the potential of machine learning, particularly deep learning models, in accurately forecasting domestic electricity demand. By integrating multiple data sources and employing advanced feature extraction techniques, the developed model significantly outperformed traditional statistical methods. The results indicate that machine learning can effectively capture the complex and nonlinear relationships inherent in household energy consumption patterns, which are often influenced by various external factors such as weather and occupancy.

The implications of accurate demand forecasting are profound. Enhanced prediction accuracy not only optimizes energy production and distribution but also contributes to cost savings for consumers and utility companies alike. By aligning electricity generation more closely with actual demand, the model helps mitigate the risks of overproduction and underproduction, which can lead to energy waste and increased greenhouse gas emissions. This aligns with the Sustainable Development Goals (SDGs), particularly in promoting affordable and clean energy while addressing climate change.

Moreover, the study highlights the importance of developing robust forecasting systems tailored to residential settings, filling a critical gap in existing research. Future work should focus on refining these models further and exploring their application in real-time energy management systems, potentially leading to more sustainable energy practices and improved grid reliability.

In summary, leveraging machine learning for electricity demand forecasting not only enhances operational efficiency but also supports broader environmental goals, paving the way for smarter and more sustainable energy management solutions.

Reference

[1] Klyuev, R. V., Morgoev, I. D., Morgoeva, A. D., Gavrina, O. A., Martyushev, N. V., Efremenkov, E. A., & Mengxu, Q. (2022). Methods of forecasting electric energy consumption: A literature review. *Energies*, *15*(23), 8919.

[2] Kiprijanovska, I., Stankoski, S., Ilievski, I., Jovanovski, S., Gams, M., & Gjoreski, H. (2020). Houseec: Day-ahead household electrical energy consumption forecasting using deep learning. *Energies*, *13*(10), 2672.

[3] Gellert, A., Fiore, U., Florea, A., Chis, R., & Palmieri, F. (2022). Forecasting electricity consumption and production in smart homes through statistical methods. *Sustainable Cities and Society*, *76*, 103426.

[4] Lazzari, F., Mor, G., Cipriano, J., Gabaldon, E., Grillone, B., Chemisana, D., & Solsona, F. (2022). User behaviour models to forecast electricity consumption of residential customers based on smart metering data. *Energy Reports*, *8*, 3680-3691.

[5] Haq, E. U., Lyu, X., Jia, Y., Hua, M., & Ahmad, F. (2020). Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach. *Energy Reports*, *6*, 1099-1105.

[6] Hotz, N. (2024, April 28). *What is CRISP DM?*. Data Science Process Alliance. https://www.datascience-pm.com/crisp-dm-2/#How_Popular_is_CRISP-DM

[7] United Nations. (n.d.). *The 17 goals | sustainable development*. United Nations. https://sdgs.un.org/goals