Predicting future electricity usage from available data

Predicting future electricity

Sara Kfouri Koskinen

Master’s Thesis
May 2025
Master’s Degree Programme in Artificial Intelligence and Data Analytics

Jyväskylän ammattikorkeakoulu
University of Applied Sciences

Kfouri Koskinen, Sara

Predicting Future Electricity Usage from Available Data

Jyväskylä: JAMK University of Applied Sciences, May 2025, 79 pages. Master's thesis

Master's Degree Programme in Artificial Intelligence and Data Analytics

Permission for open access publication: Yes

Language of publication: English

Abstract

This thesis presents the development of a machine learning predictive model for forecasting electricity consumption in residential units in Finland, including traditional houses and seasonal summer cottages. The primary objective was to help power plants and distributors in accurately anticipate the energy demand in the future, contributing to more efficient energy management and planning.

This project focused on predicting electricity usage through data preparation and the development of a machine-learning prediction model for future energy consumption in Finland. The objective was to develop a model that accurately predicts household electricity consumption. To create this prediction model, historical energy consumption data were used, which were combined with meteorological datasets from Jyväskylä, Pori, and Vesanto. These datasets were incorporated with important climate-related factors, such as temperature, precipitation, and wind speed, into the prediction model. Additionally, data from all holiday days were integrated to add variations in energy usage patterns during these special dates to the model.

To ensure data consistency and relevance, the datasets were cleaned, pre-processed and missing values were handled. Household energy data was merged with weather data as well as holiday data based on the day. This predictive model aimed to assist both power plants and distribution companies in the future. It attempted to increase the accuracy in predicting energy consumption, contributing to more sustainable and adaptive strategies. As a result, planning and anticipating energy demand in homes would lead to more efficient management of electrical resources. This could contribute to an optimized allocation of resources, which could help in a smarter and more sustainable use of electricity.

Keywords

Data Preparation, Forecast, Energy Consumption, Artificial Intelligence, Python

Miscellaneous (Confidential information)

-

Acknowledgment

I would like to express my deepest gratitude to my husband for his unwavering support and help during this crucial time when I needed to dedicate time away from family to work on my master's degree.

I also want to thank my son, Arto, for being such a wonderful boy, understanding my need to work, and patiently playing on his own when I was busy.

Additionally, I extend my gratitude to my thesis supervisor and, especially, to Tarja for always being kind, supportive, and understanding throughout this journey, even when it took longer than expected.

I am grateful to Mika Rantonen for giving me the opportunity to carry out this thesis and for all the support throughout my studies.

Lastly, I would like to thank the company Behavior Moves, and in special Jari Parkkisenniemi and Pasi Valoranta, for the inspiring idea behind this thesis and for their continuous support along the way.

Contents

Figures

Tables

Terms and Abbreviations

AI	Artificial Intelligence
DB	Database
DGP	Data-Generating Process
DP	Data Preparation
DQ	Data Quality
LSTM	Long Short-Term Memory
LGBM	(LightGBM) Light Gradient Boosting Machine
Dimension	The term dimension and data dimension is used as data feature and data attribute, something equivalent as a column in an excel sheet.
Feature	Dimensions are named “feature” with the names of the attributes of the dimension in the code.

1 Introduction

The ongoing war between Ukraine and Russia has changed many things on the world stage. It has also triggered a huge economic crisis and, above all, an energy crisis throughout Europe. After the start of the war, several nations realised how dependent they were on electricity. These countries are just consumers and not producers. Finland, as well as other European nations, began diligently pursuing new sources and forms of energy in an attempt to move away from dependence on other countries while also seeking more self-sufficient energy strategies.

The more accurate the data is in predicting average energy consumption, the better. Knowing the average consumption of each home, business, or industry in the short, medium, and long term can improve energy planning for energy supply and distribution companies. Accurate energy demand is fundamental for companies to be able to plan production and distribution efficiently to meet regional needs. This is particularly important in countries like Finland, where energy production is often dependent on weather conditions.

According to the International Energy Agency (2024), global electricity demand grew at a slower rate in 2023 compared to 2022, resulting in a 0.2% decline. Energy consumption growth in 2022 was 2.4%, and in 2023 it was 2.2%. The decline was mainly due to a decrease in electricity consumption in advanced economies. The decline was mainly due to a decrease in electricity consumption in advanced economies. Unfortunately, the decline did not continue; on the contrary, forecasts indicate that demand will recover in the next three years, with an average annual growth rate of 3.4% expected through 2026. The return to growth in energy consumption is directly linked to the stronger global economy, which drives electricity use in developing countries and superpowers.

In line with the International Energy Agency, emerging and developing economies are expected to account for 85% of global electricity consumption growth by 2026. China is expected to be a key player in this scenario, contributing an annual increase of 6.4% due to strong industrial activity.

China could also potentially consume the equivalent of half of the European Union's annual electricity use (International Energy Agency, 2024).

Another factor that is expected to increase electricity consumption considerably is technological advances. Artificial intelligence and data centres are expected to contribute significantly to this growing demand. Data centres are expected to double their electricity consumption, surpassing Japan’s annual use. To meet this growing demand, there will, therefore, be a need for new regulations focused on sustainability (International Energy Agency, 2024).

In addition, we have to take into account the effects of climate change, which are having a major impact on the present and future. From changes in habitats, wildlife, and temperatures to the distribution of water on Earth. In 2024, it was observed that there was an increase in the frequency of heat waves, which can reduce the ice in the Baltic Sea. There is a considerable influence on energy consumption due to the impact of climate change. Extreme temperatures are being observed in Finland. Climate change effects are changing the hydrography of the planet. In January 2024, the Finnish Meteorological Institute issued a warning of temperatures below -40°C. In the same year, in May, there was an intense heat warning issued by the Finnish Meteorological Institute (Finnish Meteorological Institute, 2025).

Given the current global context and trends in energy consumption, Finland must have a sustainable electricity system. This project intends to develop a forecasting model for residential energy consumption. By accurately predicting energy needs, the model will support energy companies to achieve more efficient and responsive energy production, ultimately improving sustainability and operational effectiveness.

1.1 Background and motivation

The aim of this project is to improve the short-term efficiency of energy production by developing a model that accurately forecasts residential energy consumption. In modern energy systems, particularly those that include renewable sources, it is absolutely essential to anticipate consumption in advance. This allows energy providers to better align production with demand, reduce waste, lower operational costs, and maintain a stable and reliable energy supply.

Predicting energy consumption at the household level has become increasingly important in the context of the European energy crisis, which exposed the vulnerability of energy systems to geopolitical tensions and market volatility, mainly after the war between Russia and Ukraine started making evident the dependency that many European countries had (International Monetary Fund, 2022).

A key feature of “consumption without production” is that goods and services are traded and used before they are produced. Consumption without production is a common practice in wholesale and retail markets, which varies according to speculations about future production. Influenced by finance in the management of production and distribution, the imbalances reveal the complex relationships between systematic imbalances in global markets and economic forces (International Energy Agency, 2024).

The increasing demand for the socialization of industrial processes and the globalization of production and distribution further intensifies this trend, especially in energy markets. “Consumption without production” acts as an informal mechanism to balance the high demand for energy commodities that drives much of economic activity, requiring a constant flow of resources (Zatzman, 2012).

Finland benefits from a mix of nuclear power and renewable energy production, particularly through forest biomass, hydropower, and wind power. As a result, Finland has one of the lowest fossil fuel dependencies among IEA member countries, regardless of challenges such as high energy demands driven by its cold climate. Finland’s transition to renewable energy supports global climate change efforts and promotes a balanced and resilient economy (International Energy Agency, 2023).

Finland’s shift towards sustainable energy underscores this historical trend of “consumption without production.” By reducing its dependence on fossil fuels, Finland strives to minimize the potential hazards linked with past economic consolidations. Finland’s change could promote a more balanced and more resilient economy (International Energy Agency, 2023).

Energy forecasting is the primary component in supporting Finland’s transition to a greener and more efficient energy system, especially as the country strives for carbon neutrality by 2035. This will be achieved by enabling the integration of variable renewable energy sources such as wind. Accurate energy forecasting improves demand management and helps stabilize the grid, and also supports dynamic pricing, reduces peak loads, and promotes sustainable energy habits among consumers. This initiative seeks to contribute to these national goals by increasing the accuracy of forecasting residential electricity consumption, contributing to improved, smarter energy planning, and more sustainable use of resources (European Commission, 2022).

While energy forecasting has been studied before, this thesis moves it forward to real-life use in Finland. It highlights the necessity to forecast the immediate energy consumption of each individual residence to determine how much energy should be produced. By using hourly data from individual homes, combined with local weather and holiday information, and seasonal demand, making their planning more precise and effective energy planning.

The project also contributes to sustainability by promoting more efficient energy consumption at the household level. Allowing for more accurate forecasts, the model supports smarter energy management, which helps reduce unnecessary energy production and avoid peak-time overloads. This not only reduces emissions but also supports more stable energy systems. Furthermore, the ability to anticipate consumption patterns empowers both consumers and energy providers to make well-considered choices, conforming to the goals of sustainable development and improving corporate awareness.

1.2 Problem statement and research questions

The recent energy crisis in Europe, intensified by the war in Ukraine, has underscored the urgent necessity for countries to become more energy self-sufficient. The disruption in supply chains and rising energy prices have put significant pressure on governments and energy providers to providers more efficient and sustainable solutions. In this context, the capacity to precisely forecast residential energy consumption, especially in the short term, has become vital for efficient planning and cost reduction.

This study analyses the importance of short-term energy consumption prediction to improve market efficiency, reduce costs, and optimize supply. The investigation is based on two main questions:

1. **Can energy consumption be predicted for the next 24 hours on an hourly basis?**

This question highlights the importance of hourly energy consumption forecasting for price improvement in markets that use day-ahead trading for pricing, but also for enabling short-term operational planning and maintenance by providing accurate predictions a few hours in advance. Short-term forecasting improves accuracy and reduces financial risks by overestimating or underestimating production and consumption, thus leading to better economic outcomes for all (Barthelmie, 2008).

2. **Can energy consumption be predicted for the next five days on an hourly basis?**

Extending the prediction to five days with average hourly consumption allows for more comprehensive planning. This also optimizes energy storage and prepares for climate impacts. Preparing for climate impacts is mainly important in Finland, where winter often presents very low temperatures and consequently increases energy consumption (Essenfelder et al., 2020).

An accurate energy consumption prediction can facilitate the decision-making process, resulting in greater energy efficiency and, consequently, lower costs (Suganthi & Samuel, 2012).

By addressing these questions, this research seeks to develop predictive models that improve the company’s ability to dynamically respond to changes, ultimately contributing to a more efficient energy demand.

In order to effectively predict home energy consumption, mainly during times of significant demand unpredictability, the companies want to develop predictive algorithms. Production and supply can be optimized by anticipating and recognizing consumption trends. In addition to helping with strategic planning, these insights may also lower operational costs and consequently the final price for the citizens (Barthelmi et al., 2008).

1.3 Research objectives

The main objective of this project is to create a model to predict electricity usage in Finland through data preparation, as well as to create a machine learning forecasting model to accurately predict future energy consumption. The goal is to create predictive models that forecasting residential energy consumption in the short term (next 24 hours) and (next five days), with hourly forecasts.

By offering more accurate forecasts, the model helps energy providers plan and produce electricity more efficiently. This means they can better match production with actual household consumption, which helps avoid unnecessary overproduction and reduces energy waste. It also contributes to a more stable and reliable energy supply for everyday use. For power plant companies, having access to dependable short-term forecasts supports smarter decision-making. With this information, they can customize their operations, manage demand more effectively, reduce costs, and align more closely with their long-term goals (Barthelmie et al., 2008).

Companies could offer predictive models to power plants companies to help them forecast short-term energy demand more accurately. With better forecasts, power plants can optimize their production process, reduce waste, and operate more efficiently.

Energy consumption has become fundamental to modern life, so it is indispensable to understand the need for truly efficient planning. The increase in demand and increase variability in consumption have made accurate forecasting mechanisms essential. For example, the use of predictive models to optimize energy distribution and production especially after the war in Ukraine, which highlighted Europe’s vulnerability to energy supply and distribution emphasized the importance of this project.

The project develops a prediction model for energy consumption of residential and summer houses. Based on historical data and the trend of high consumption during holidays, the model will focus on delivering accurate estimates of future consumption, allowing for more effective planning. By accurately predicting energy needs, the model will be able to support companies in the sector to achieve more efficient and responsive production. This will contribute to improving sustainability and functional productivity, reducing waste, and optimizing resources.

This research-based development project focuses on building and evaluating machine learning models for forecasting residential energy consumption. The project was commissioned by the consultant company Behaviour Moves. The primary goal is to develop models that can accurately predict short-term energy consumption using historical consumption data and weather information.

Although electricity consumption forecasting has been previously studied, this thesis introduces an advanced approach by applying Transformer-based models to short-term hourly predictions for each household, representing a new application of this technology in energy forecasting. It integrates weather and calendar data specifically tailored to Finnish homes. In addition, two preprocessing approaches were implemented to align with the specific characteristics of each algorithm. The result is a practical model with real commercial potential for power plant operators, supporting more effective daily energy planning.

1.4 Scope and limitations

This study focuses on forecasting hourly residential energy consumption over two different time horizons: the next 24 hours and the next five days. The forecasting models are developed based on historical energy consumption and weather data from six households located in Finland.

The dataset only includes households with fixed-price electricity contracts, meaning that consumption behaviour is not directly influenced by hourly fluctuations in market prices.

The study is limited to short-term forecasting. It does not include long-term forecasting tasks, demand response optimization, or the integration of additional external factors such as occupancy data or heater types.

1.5 Research ethics and data handling

This study follows best research practices and complies with the ethical guidelines for research at JAMK. The quantitative project focuses on time series forecasting using regression-based machine learning models.

The datasets used in this thesis consist of energy consumption records (anonymized) and public meteorological data. No personally identifiable information (PII) was collected or processed. Each household was assigned a random house ID to ensure anonymity.

The data collection and processing procedures complied with the principles of the General Data Protection Regulation (GDPR) (EU 2016/679), which provided the energy data, confirming that the data had been anonymised before sharing.

No human participants were directly involved in the research, and no interventions were performed. The research did not require separate ethical review by an ethics committee because it did not involve sensitive personal data, health-related data, or vulnerable groups.

By following these ethical principles, the study ensures that data privacy, confidentiality, and the responsible handling of information are respected throughout the research process.

This research follows a quantitative approach, using machine learning regression models to predict future energy consumption based on historical time series data.

Quantitative research focuses on measurable data and statistical analysis. This thesis used historical energy consumption and meteorological data to build and validate predictive models. The study does not involve qualitative methods, interviews, or subjective assessments.

During the preparation of this thesis, Artificial Intelligence (AI) tools were used, specifically OpenAI's ChatGPT (released March 14). The tools were used to assist in brainstorming ideas, research, improve the clarity of some technical definitions, find references, and revise the structure of certain sections.

However, AI was not used to generate research results, provide writing support, or assist with language refinement, write essential analyses, or replace the author's critical thinking. All use of AI was carefully reviewed, edited, and supplemented by the author to ensure academic integrity and originality. Final responsibility for the content lies solely with the author.

1.6 Project structure

Once the project topics was defined, the theoretical research phase began. The first step involved identifying the most commonly used technologies for build investigation started with the search for the most widely used technologies for creating energy forecasting models. The technologies chosen through the theoretical study were presented in the second chapter. The models selected for this study were Long Short-Term Memory (LSTM) networks, Light Gradient Boosting Machine (LightGBM), and the Head Attention Transformer model. These models were chosen due to their ability to handle complex patterns in time series data and their proven success in previous research on forecasting.

Python was chosen as the primary programming environment due to its flexibility and strong support for data science applications. The advantages of using Python are numerous, starting with the simplicity of its structure and ease of use. It can also be used in data analysis. Python is widely used in the development of real-world implementations that use Artificial Intelligence in machine learning models, such as this electricity consumption forecasting model (Arshad, M., 2024).

Pandas was used for data manipulation and preprocessing. Matplotlib was used to create visualizations for data exploration and results analysis. TensorFlow was used to create the LSTM sequence model and the Head-Attention Transformer model. The selection of these tools ensured an efficient development process aligned with the project objectives (Eichner, A., 2023).

All data used for model development are covered in chapters three and four. Details of data description, meteorological data, data privacy, and moral principles were explored in chapter three. Chapter four explores the data preparation process, including data collection, preprocessing workflows, merging of datasets, and an assessment of data quality.

The fifth chapter discusses the three energy consumption prediction models developed: the LSTM-based model, the transformer-based model, and the LightGBM model. In the sixth chapter, all models were visualized, evaluated, and compared. The last two chapters were dedicated to the final discussions, strengths, weaknesses, recommendations, future development, conclusion, and final thoughts.

1.7 Research methods

This study adopts a quantitative research approach to predict electricity consumption in residential environments using machine learning techniques. The methodology is based on the analysis of historical data that has been structured in time series, which allows the identification of patterns and trends in energy consumption. In the case of this study, electricity consumption takes into account meteorological variables relevant to energy consumption, as well as temporal indicators that influence energy consumption, such as identifying whether the data refers to weekdays or weekends and holidays (Hyndman & Athanasopoulos, 2018).

The objective is to develop energy consumption predictive models capable of recognising complex relationships in the data to produce accurate short and medium-term forecasts.

To build these models, two main categories of explanatory variables were considered. The first category includes meteorological variables such as temperature, humidity, precipitation, and wind speed, which directly influence the use of heating systems, which is the largest influencer of residential energy consumption. The second category consists of temporal variables, including time of day, day of the week, national holidays, and the distinction between weekdays and weekends. These variables help characterise users' routines and their impact on energy demand. Combining these two types of data allows models to learn how external environmental conditions and temporal factors jointly influence residential electricity consumption.

The data used in this research were collected from a specific set of households. However, the study was not conducted as a traditional case study. Instead, the research was designed to produce generalizable results that could be applied to other residential contexts, provided similar input variables and data infrastructure are used. This broader perspective increases the practical applicability of the study's results across different regions or systems with similar characteristics.

Three machine learning architectures were implemented, evaluated, and compared to assess their suitability for time series forecasting in the energy domain. These prediction models were selected based on the literature that claims and proves their ability to handle sequential data and discover non-linear patterns working with complex datasets.

The first model is based on Long Short-Term Memory (LSTM) networks, enhanced with attention mechanisms to help the model focus on the most relevant parts of the consumption sequence.

The second model follows the Transformer architecture, originally developed for natural language processing (NLP). Long Short-Term Memory Model (LSTM), this architecture was used because it demonstrated strong performance in capturing dependencies in time series data in prediction models (Vaswani et al., 2017).

The third model created uses gradient boosting via the LightGBM library, which was selected because of its great performance for processing large datasets and multiple input variables for forecast models (Ke et al., 2017).

The project data were preprocessed to ensure quality and consistency before starting the model training. This is the best practice, including handling missing values, identifying and correcting outliers, and also normalising numerical variables. Feature engineering played a central role in this step, extracting important features such as temporal indicators and weather variables that have a strong influence on household energy consumption (Zhang et al., 2018).

The temporal structure and the chronological integrity of the dataset were preserved during training and evaluation, providing a more reliable assessment of the models' predictive ability. The performance of each model was evaluated using widely accepted metrics. These metrics allow a comparative analysis of the prediction accuracy. The quantitative research of the project is based on the analysis of the three models' results. The following metrics were used: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2).

In summary, the methodological design adopted in this research allows a quantitative evaluation of machine learning models for predicting residential energy consumption, which can corroborate their potential integration into broader energy planning and management strategies.

2 Literature review and theoretical background

2.1 Energy forecasting and its importance

Accurate energy consumption forecasts are very useful for ensuring the efficiency, reliability, stability and sustainability of electricity supply. In the future, they will facilitate energy suppliers' planning, leading to the optimisation of energy production, reduction of operating costs and thus avoiding possible blackouts or even overproduction. In Finland, especially where electricity consumption patterns are highly influenced by climate interactions, the more accurate consumption forecasting strategies are, the better. Forecasting at the household level enables more accurate demand management, helping consumers and suppliers to respond proactively to supply and price fluctuations (Doumèche et al., 2023).

Overall, energy forecasting is not only a technical challenge, but also a key enabler for building smarter and more sustainable energy systems. Forecasting at this level of detail enables consumers to make informed consumption choices, contributing to cost savings and to the broader effort to strengthen Europe's energy resilience and sustainability (Enerdata, 2024).

During periods of energy shortages or sharp price increases, as occurred in Europe following the reduction in natural gas imports from Russia in 2022 due to the war, the role of energy consumption forecasting becomes even more strategic. Accurate forecasts allow utilities, grid operators, and policymakers to better understand where, when, and how energy is being consumed (European Commission, 2022).

The models selected in this thesis were guided by existing research demonstrating their demonstrating their accuracy in time series forecasting tasks. For example, Munir et al. (2023) showed that LightGBM demonstrates reliable performance on structured datasets, making it a suitable choice for energy consumption prediction.

Similarly, the LSTM was selected for its strength in identifying patterns that develop over extended periods in sequential data, as demonstrated by Siami-Namini et al. (2019) in various forecasting applications.

The inclusion of the Transformer model was influenced by the findings of Wu et al. (2021), which showed that it performs well in handling long-range temporal dependencies compared to traditional recurrent models. These studies provide a foundation for evaluating the selected models and provide a useful reference for comparing the patterns observed in the thesis.

2.2 Time series forecasting concepts

A time series is defined as a sequence of data points collected at successive intervals over time. The main characteristics of a time series are its temporal order, which is what distinguishes it from other data structures. By analyzing a time series, patterns such as trends, seasonality, and cyclical behaviour can be identified, making it a valuable tool for predicting future observations. For example, recording hourly energy consumption, daily stock prices, or hourly weather conditions generates time series data. These datasets enable research and analysis where analysts can extract meaningful insights and predict future values based primarily on historical trends (Joseph, 2022).

This thesis explores various methods for pre-process data for time series analysis and examines three supervised learning models, with a specific focus on time series forecasting. The initial models tested include attention-based models (Head-Attention), LSMT, and LightGBM regression learning approaches suitable for capturing complex temporal patterns, such as energy consumption. Each of these models has distinct characteristics that make it suitable for different aspects of time-series prediction. (Li, 2024).

Vaswani et al (2017) advises the use of attention-based models, specifically the Multi-Head Attention Mechanism, which applies a deep learning technique that was developed to deal with high-dimensional data, as is the case with this prediction model. In prediction applications that allow models to process complex and multivariate data through information partitioning, multiple factors can influence the outcome (Wang et al., 2023).

Energy consumption and weather data are typically recorded at uniform intervals. This thesis focuses on a regular time series to estimate future energy usage based on historical consumption data and external factors, such as holidays, weather conditions and previous hour consumption.

The historical consumption guided the thesis to use Supervised Learning, which is based on the concepts of "labelling," involves a system using the input data (independent variables/features/temperature) and corresponding labels (dependent variables/target/real value consumption of energy), allowing it to forecast future values (Glassner, 2021).

In the project task, the model learns to associate real consumption values with corresponding features, which allows it to predict labels accurately on new data.

2.3 Machine learning in forecasting

Machine learning (ML) methods are an effective tool for forecasting energy consumption by identifying complex patterns in large volumes of data. Machine learning has been widely applied to forecasting time series, such as energy demand. These models enable short- and long-term forecasts, facilitating more efficient distribution, dynamic pricing, and integration of renewable sources. In residential consumption, ML models use historical data combined with variables such as weather and holidays to increase the accuracy of forecasts.

Machine Learning methods have become important in time series forecasting. One important ability that the Machine Learning models have is to work with complex, non-linear relationships. Compared to traditional statistical techniques, machine learning models to better adapted to fluctuations and patterns in energy consumption data, particularly when influenced by multiple external factors such as temperature and holidays.

2.4 Recurrent Neural Networks and LSTM

Recurrent Neural Networks (RNNs) are widely used in tasks that involve sequences, such as time series forecasting. While they can process sequential data, traditional RNNs often face difficulties when working with long sequences due to the vanishing gradient problem. This makes it harder for them to retain important information over time. To solve this LSTM networks were developed. Thanks to their internal gating structure, LSTMs can decide what information to keep or discard across different time steps, which helps them handle long-term patterns more effectively (Geff et al., 2017).

LSTMs are particularly effective in energy consumption forecasting due to their ability to model temporal patterns over time. They can retain memory of past data points, which helps capture seasonal changes or repeating daily and weekly trends in consumption. Additionally, they handle non-linear relationships, making them suitable for energy datasets influenced by weather, time, and behavioural patterns.

LSTM networks can also support multi-step forecasting. They can be implemented in a sequence-to-sequence structure, where the model outputs a series of predictions instead of just one, making them ideal for both 24-hour and multi-day forecasts. However, they require large amounts of data to generalize effectively and are computationally expensive, especially when training on long sequences. The key advantages of LSTM models include memory retention, the ability to model long-term dependencies, capture non-linear temporal relationships, handle variable-length sequences, and perform multi-step forecasting effectively (Brownlee, 2020).

One disadvantage of LSTM models is the requirement for a large dataset for effective generalization and can be computationally expensive (Bengio et al., 1994).

2.5 Transformer models in time series

Transformer-based architectures have recently been applied to time-series forecasting and regression tasks. Unlike traditional models such as Long Short-Term Memory (LSTM) networks, Transformers use attention mechanisms to evaluate the relevance of previous data points, allowing them to model temporal relationships without relying on sequential processing (Zerveas et al., 2021).

Transformers have shown strong performance in time-series regression, where the goal is to predict a continuous value (e.g., electricity consumption, stock prices, weather conditions) based on historical data (Wu et al., 2021).

The key advances of Transformers models for energy consumption forecasting include the use of self-attention to capture long-range dependencies by processing all time steps simultaneously. The self-attention mechanism dynamically determines which past time points are most relevant for making future predictions. Transformers also process all time sequences in parallel, which drastically reduces training time and makes them especially beneficial for large time series datasets. Additionally, they can integrate multi-resolution information, improving prediction accuracy across different time horizons (Oliveira & Oliveira, 2023).

Transformer-based models are known for their high computational cost and typically require large datasets for effective performance. Furthermore, although they are powerful for sequence modelling, they do not capture temporal dependencies like traditional time series models. In fact, transformer-based models rely on positional encodings to incorporate sequence order, which may not always fully capture the temporal structure of energy consumption data (Vaswani et al., 2017).

2.6 Data generating process and regression analysis

Time series data originates from the underlying mechanism referred to as the Data Generating Process (DGP). In energy consumption forecasting, the DGP is affected by variables such as climate conditions, economic activity, and consumer behaviour. Understanding the DGP is essential for building accurate forecasting models to approximate the DGP by identifying statistical relationships within historical data (Joseph, 2022).

In energy consumption forecasting, the Data-Generating Process (DGP) is affected by variables such as climate conditions, economic activity, and consumer behaviour. Understanding the DGP is essential for building accurate forecasting models, as it provides insights into the factors influencing energy consumption patterns.

For Joseph (2022), from a theoretical perspective, DGP (Data-Generating Process) is often represented as a stochastic process, incorporating random fluctuations that make accurate predictions very challenging. Instead, machine learning models approximate DGP by identifying statistical relationships within historical data.

From a practical perspective, DGP is often represented as a stochastic process, incorporating expected fluctuations or variations that make accurate predictions challenging. Instead, machine learning models approximate DGP by recognising statistical relationships in past data (Joseph, 2022).

The diagram below (Figure 1) illustrates the process of time series forecasting using machine learning. External factors (such as weather variations and holidays) influence the DGP, which produces time series data. A machine learning model can learn from historical data to approximate the DGP and generate forecasts for future values.

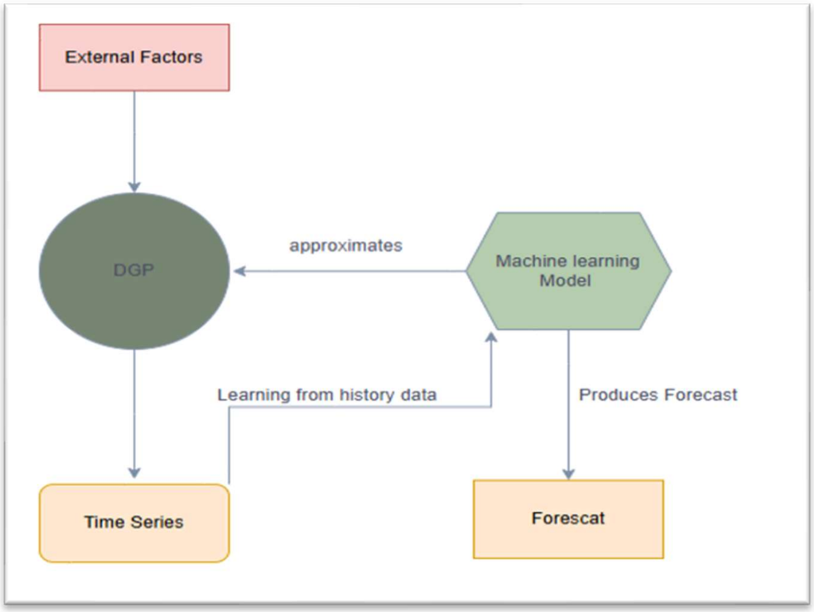


Figure 1. Process of time series forecasting using machine learning

Several forecasting approaches exist, including traditional regression models and advanced deep learning techniques such as LSTM, Attention-based models, and LightGBM.

Regression analysis enables the dependent variable (e.g., energy consumption) to be estimated using one or more independent variables (e.g., weather, day, holidays) (Tabasi & Forotan, 2016).

Regression analysis is an estimation analysis method that enables the dependent variable to be estimated with the help of independent variables. The state of dependent variables can be defined using regression analysis according to one or more independent variables. This relation between these variables can be represented through a regression model (Tabasi & Forotan, 2016).

For instance, if Y represents energy consumption and X represents influencing factors such as weather, day, holidays (denoted as $X_1, X_2, X_3, \dots X_n$). The relationship can be defined as:

$$Yi = \beta 0 + \beta 1\, X\{1i\} + \beta 2\, X\{2i\} + \dots + \beta n\, X\{ni\} + \varepsilon i. \tag{1}$$

Where:

- $\beta 0, \beta 1, \dots, \beta n$ are regression coefficients estimated to minimise error,
- εi is the residual error term.
- The optimisation of these coefficients ensures the best possible estimation of Y (Fumo & Biswas, 2015).

Regression models can be linear or nonlinear, depending on the relationship between variables.

2.7 Gradient boosting and LightGBM

Gradient boosting models, such as Light Gradient Boosting Machine (LightGBM), are frameworks for machine learning models widely used for time series forecasting, in special with structured tabular data (Ke et al., 2017). Unlike deep learning models that require extensive feature engineering, tree-based models naturally handle missing data, categorical variables, and outliers while being computationally efficient (Shwartz-Ziv & Armon, 2022).

Florek & Zagdański (2023a) point out that the tree growth employed by LightGBM is different from traditional augmentation methods since the framework uses a leaf-by-leaf tree growth strategy. In this case, LightGBM always selects the leaf with the greatest loss reduction, causing the trees that use it to grow deeper, which increases the improvement in data accuracy. Another advantage that Florek & Zagdański (2023b) highlight is LightGBM's ability to learn using histograms. This causes the values of continuous features to be discretised into compartments, thus reducing computational complexity and memory usage.

According to Ke, LightGBM is a gradient boosting framework designed for high efficiency and scalability, which makes it ideal for the prediction model proposed here. LightGBM is suitable because it is efficient in machine learning tasks, especially for large data sets and high-dimensional data like that needed for energy consumption prediction (Ke et al., 2017).

Advantages of LightGBM for energy consumption forecasting:

- Efficiency and Scalability: Ideal for large-scale datasets (Ke et al., 2017).
- Leaf-Wise Tree Growth: Selects the leaf with the greatest loss reduction, improving accuracy (Florek & Zagdański, 2023a).
  Histogram-Based Splitting: Reduces computational complexity and memory usage.
- Gradient-Based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB): Enhance speed without sacrificing accuracy.

In addition, the Light Gradient Boosting Machine (LightGBM) algorithm improves training efficiency through two key techniques. The first, Gradient-Based One-Side Sampling (GOSS), prioritizes larger gradient values to accelerate learning without reducing accuracy. The second, Exclusive Feature Bundling (EFB), helps reduce the number of features by combining those that do not overlap, simplifying the model and improving speed. These features make LightGBM faster and often more effective than other commonly used frameworks, such as XGBoost (Florek & Zagdański, 2023a).

For Munir et al. (2023), LightGBM is well suited for a variety of data-driven tasks, particularly those involving large or sparse datasets. It also supports parallel computing and GPU acceleration, which enhances performance on complex problems. This makes it a practical choice for forecasting projects like this one, which involves detecting seasonal patterns and responding to changing weather conditions.

Previous studies have demonstrated the practical value of machine learning models for energy consumption forecasting. Wu et al. (2021) showed that Transformer-based architectures outperform traditional LSTM networks in long-range forecasting tasks. Munir et al. (2023) found that LightGBM models achieve strong predictive performance on structured datasets, such as residential energy consumption.

2.8 Model evaluation metrics

In time series forecasting, evaluating model performance is crucial to ensuring accurate and more reliable forecasts. Several statistical metrics can help measure the reliability of these forecasts and how well a model's predictions match the actual values. The Jedox blog explains the use and differences between these metrics in forecasting models. A commonly used metric to evaluate performance is the Mean Absolute Error (MAE), which calculates the average absolute difference between the predicted values and the observed values. MAE provides an intuitive measure of accuracy, in which case, lower values indicate better forecasts.

Another important metric for evaluating model performance is the Root Mean Squared Error, or RMSE. Unlike the MAE, which only calculates the absolute difference, the RMSE gives more weight to larger errors by squaring the differences before calculating them. This makes the RMSE useful in cases where significant forecast errors need to be penalised more crudely (Jedox blog).

The coefficient of determination (R^2) was employed to assess the model's predictive accuracy, indicating the proportion of variance in the dependent variable explained by the model (Investopedia, 2009).

2.9 Comparison models

LSTM, Transformer, and LightGBM models were chosen based on their ability to handle time series data, although their strengths and limitations differ. This thesis compares these three forecasting models, which have been frequently used in time series prediction. This section summarises those differences and explains the reasons behind their inclusion in this project.

LSTM networks have been used successfully in many sequential data problems. As demonstrated by Siami-Namini et al. (2019), LSTM performs well when forecasting time series with trends or seasonal patterns. Its design allows the model to “remember” important information from previous time steps, which is helpful in energy forecasting where past patterns influence future use. However, training LSTM require considerable time, and the model can become unstable if the dataset is small.

Transformers take a different approach compared to traditional neural network models like LSTM. Instead of processing one time step at a time, they can handle all time steps simultaneously using a self-attention mechanism. This allows the model to focus on the most relevant parts of the sequence, regardless of their position. Wu et al. (2021) showed that the ability of Transformers to process time steps in parallel makes them particularly effective for long-range forecasting tasks, especially when large datasets are available. Because Transformers do not depend on sequential processing, Transformers tend to train faster than LSTM models. However, they require more computational resources and may not perform as well when the dataset is small.

LightGBM, unlike neural network-based models, relies on a different structure. It is a gradient boosting algorithm specifically designed for structured, tabular data. As noted by Munir et al. (2023), LightGBM is fast, handles missing values effectively, and performs well when the data includes time-related features such as lag variables. It is also easier to interpret compared to deep learning models. However, since LightGBM does not inherently recognize temporal sequences like LSTM or Transformer models do, it requires additional feature engineering to capture time-based patterns.

While LSTM is useful when sequence order and long-term memory are essential. Transformer model is powerful when working with large datasets that involve complex time patterns, and LightGBM offers a faster and more accessible alternative. These models in this thesis allows for a valuable comparison of different forecasting strategies and offers a deeper understanding into how each model performs in the real-world context of energy consumption forecasting. Together, these models support the evaluation needed to address the research objectives and questions.

3 Data sources and description

3.1 Overview of the dataset used

The datasets for this study are based on energy consumption records collected from several residential properties in Finland, including traditional Finnish houses and cottages (“Mökki”). In addition, data were collected from the nearest weather stations in Jyväskylä, Pori and Vesanto, where the houses are located. The collected data on the energy consumption of the houses were kindly provided by the consulting firm Behaviour Moves. The data were used with due respect for the anonymity of the data subjects.

3.2 Residential energy consumption data

The energy consumption dataset includes records from six households in Finland from 2017 to 2024. The dataset is structured at an hourly resolution (PT1H) and includes energy consumption values in kilowatt-hours (kWh). In addition, the dataset contains information about the time (day, month, year, and hour) corresponding to consumption values. These houses represent typical households with various occupancy patterns, family sizes, and daily habits (not specified in the dataset of the habits).

In addition to standard residences, data were collected from a traditional Finnish “mökki” or cottage. These seasonal retreats, often located in rural areas, provide insights into energy usage patterns in leisure-oriented secondary residences.

All the houses dataset has the same columns Resoluutio (Resolution), Yksikkötyypi (Unit of Measurements), Lukeman Tyyppi (Type of values) Alkuaika (timestamp), Määra (The amount of the energy consumption) and Laatu (quality of the data) the Table 1 – Residential dataset Description.

Table 1. Residential Dataset Description

Column	**Description**
Resoluutio	All rows from all the datasets are in PT1H format, indicate hourly measurements.
Yksikkötyyppi	All rows from all the datasets are in kWh format, indicate kilowatt-hours.
Lukeman tyyppi	All rows from all the datasets in BN01, indicate that the contract is based on a fixed price.
Alkuaika	The day, month, year and hour of each measurement.
Määrä	Represents the total consumption per hour as indicated by Yksikkötyypi.
Laatu	This indicates the quality of the data. In the datasets, there are only two values “Ok” and “EST” (estimated).

To ensure that each home's energy usage patterns were accurately identified, each dataset was assigned a unique home ID. This unique ID played a key role in making accurate predictions for individual homes and also served to preserve the anonymity of homeowners. Including this feature allows the prediction models to better recognise the home-specific usage trends, leading to more accurate predictions.

While these pre-processing steps have been applied to all models, the next stage of data preparation may differ depending on the type of model being used. Using neural networks and decision tree models requires different transformations and optimisations, which will be addressed separately moving forward.

3.3 Weather data and external factors: meteorological data

Meteorological data from Jyväskylä, Pori and Vesanto were integrated into the analysis to link the energy consumption patterns with the weather conditions. This approach might allow the examination of regional variations in energy demand influenced by changing climate factors. All the columns are based on the weather conditions recorded by hour in Table 2, Meteorological data.

Table 2. Meteorological Data

Column	**Description**
Observation station	This refers to the location of the weather station. Example Pori
Year	The year of the recorded observation.
Month	The month of the recorded observation
Day:	The day of the recorded observation.
Time [Local time]	The local time of the recorded observation.
Precipitation [mm]	The amount of precipitation (rain, snow, etc) in millimetres
Average air pressure [hPa]	The average atmospheric pressure recorded during the period.
Maximum gust speed [m/s]	The highest recorded wind gust speed (meter/seconds) this period.
Average wind direction	The average direction of the wind (degree) during the period.
Maximum wind speed [m/s]	The highest wind speed (meter/seconds) recorded during the period.
Wind speed [m/s]	The average wind speed (meter /seconds) during the period.
Average relative humidity [%]	The average relative humidity (percentage) during the period.
Minimum temperature [°C]	The lowest temperature (Celsius) recorded during the period.
Maximum temperature [°C]:	The highest temperature (Celsius) recorded during the period.
Average temperature [°C]	The average temperature (Celsius) during the period.
Snow depth mean [cm]	The average depth (centimetres) of snow on the ground during the period.

3.4 Data privacy and ethical considerations

Data privacy and ethical considerations were previously mentioned in the first chapter in section 1.5 Research Ethics and Data Handling. In this section, this study reaffirms its commitment to the collection, handling and privacy of the data used. Complying with the best research practices, following the ethical guidelines of JAMK University of Applied Sciences.

4 Data exploration and preprocessing

The importance of data preparation before processing any data analysis is crucial. Data Analysis and Artificial Intelligence both depend on good data preparation. Well-prepared data can be used more efficiently and on a large scale. Every reliable result of machine learning models depends on good care of the data preparation. This is the only way forecasting can be trusted.

Data preparation is one of the most important aspects of the models. Before starting any kind of analytical project, according to David Loshin, data quality should be analysed the dimensions, including accuracy, consistency, integrity, and consistency (Loshin, 2006, pp. 8-10).

The author Dejan Sarka, in his book about data preparation, comments that in any analytical project, data preparation is crucial. Even though it is the longest part of a project, and it can sometimes be very repetitive and even tedious, however, the success of a project depends deeply on data preparation. However, the success of a project depends deeply on data preparation (Sarka, 2021, p.69).

Data preparation is the most important stage and the most time-consuming, particularly for data analytics or AI. Many tools can help automate your data preparation, and those tools are improving day by day. However, until now, the best practice is still the hardest one. The entire data preparation process should be done mostly manually (Oslon, 2003, pp. 24-42).

Handling the data incorrectly and using bad data can give wrong answers. Wrong solutions can guide you in the wrong direction. Olson, in his book on data quality, also mentions the cost of bad data to corporations. Wrong decisions are often the focus of the problems. In reality, no one associates the wrong decisions with the bad data that led them to make wrong decisions. Usually, the bad data quality is not even identified (Olson, 2003, p. 106).

The hourly energy consumption datasets provided included the following columns: Resoluutio (Resolution), *Yksikkötyyppi* (Unit Type), *Lukeman tyyppi* (Contract Type), *Alkuaika* (Time), *Määrä* (Consumption), and *Laatu* (Quality). The *Alkuaika* column was converted to a datetime format to ensure data consistency.

Weather data from the weather stations closest to each residence were added every six months due to site limitations that restricted the time range or number of observations per dataset. These weather datasets included several columns such as Observation Station, Year, Month, Day, Time [Local Time], Precipitation [mm], Mean Air Pressure [hPa], Maximum Gust Speed [m/s], Mean Wind Direction [°], Maximum Wind Speed [m/s], Wind Speed [m/s], Mean Relative Humidity [%], Minimum Temperature [°C], Maximum Temperature [°C], Mean Temperature [°C], and Mean Snow Depth [cm].

In some columns, there were values with a "-" symbol. According to the information on the Meteorological Institute's webpage, this value means that it was not possible to measure at that time, because of that the value was replaced with null values. In the Figure 2 below the image represents the steps of the Preprocessing Workflow:

Figure 2. Preprocessing workflow for LightGBM (left) and LSTM/Transformer (right). The diagram illustrates how weather and house datasets are merged, selected, and transformed based on the model type. LightGBM requires minimal preprocessing, while LSTM and Transformer models involve scaling, encoding, and sequence formatting.

In this thesis, two distinct modelling strategies were explored: tree-based models (LightGBM and Gradient Boosting) and deep learning models (LSTM and Transformer). Although no formal pipeline frameworks were used, preprocessing was carried out through a series of Python functions and scripts designed to meet the specific input requirements of each model type. The figure below visually summarises the steps followed in both approaches.

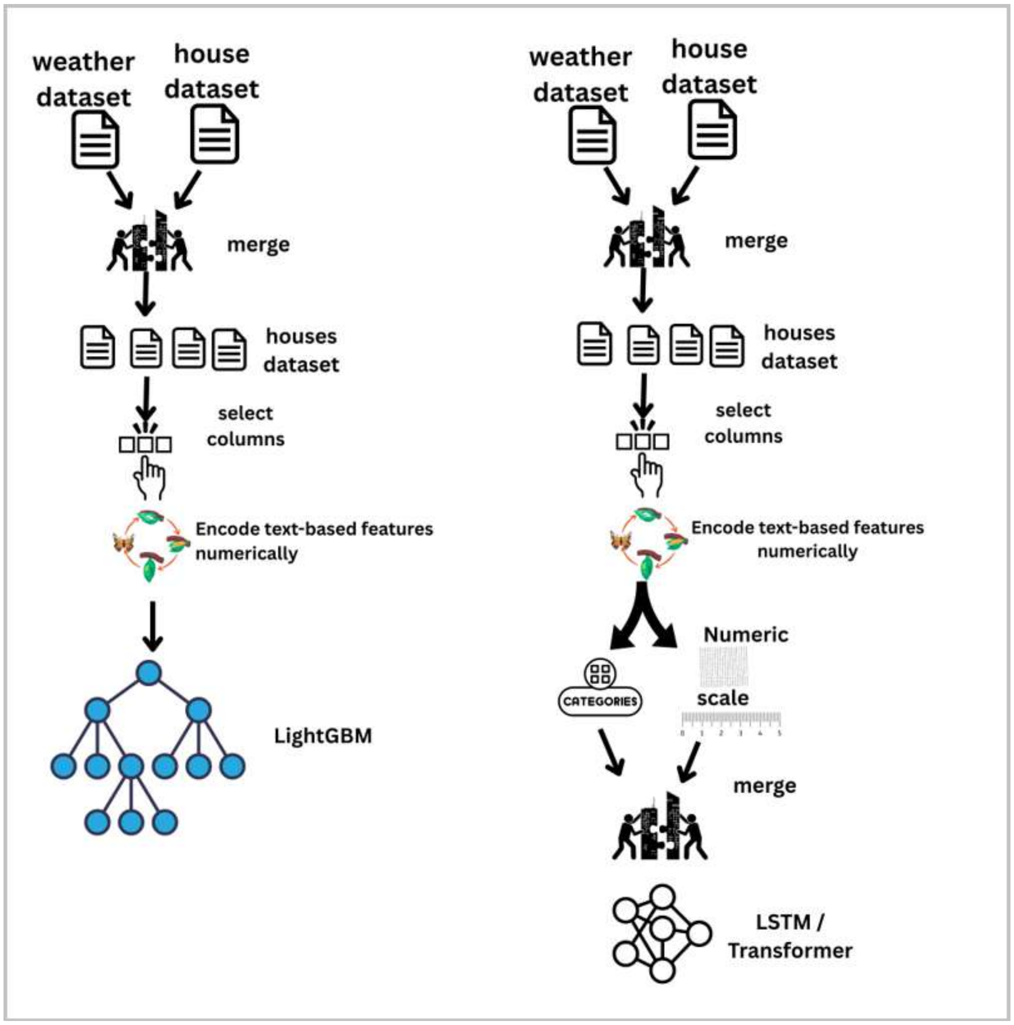


Figure 2. Preprocessing workflow for LightGBM (left) and LSTM/Transformer (right). The diagram illustrates how weather and house datasets are merged, selected, and transformed based on the model type. LightGBM requires minimal preprocessing, while LSTM and Transformer models involve scaling, encoding, and sequence formatting.

Figure 2 provides an overview of the data preparation process used for each of the modelling strategies. This diagram shows the main steps involved in transforming the original datasets into a format suitable for training the models.

On the left side of Figure 2, the diagram outlines the preprocessing steps used for the tree-based models. The weather and house datasets were first merged using the timestamp and house identifier. Following this, only the relevant columns were retained. The data was then split into numerical and categorical features. Categorical variables were encoded using either label encoding or single-value encoding, depending on the nature of the feature. Scaling was not applied to the numerical features, as models like LightGBM and Gradient Boosting do not rely on scaled inputs. These algorithms are capable of processing unscaled numerical data efficiently and are not sensitive to the magnitude of the input values (Géron, 2019).

The right side of Figure 2 illustrates the preprocessing applied to deep learning models such as LSTM and Transformer. These models required additional preparation steps due to their sensitivity to feature scale and their reliance on sequential data. The weather and housing datasets were also merged and cleaned using a similar method. However, the numerical features were scaled using normalisation or standardisation techniques. Scaling was necessary because deep learning models use gradient-based optimisation algorithms that can be affected by varying input scales. Categorical features were encoded using one-hot encoding, and the final dataset was structured into sequences to preserve temporal patterns essential for models that learn from time-based inputs (Goodfellow, Bengio, & Courville, 2016).

These preprocessing strategies reflect the practical differences between the models used in this study. Tree-based models are more flexible with input formats and typically require minimal preprocessing. In contrast, deep learning models depend on well-prepared input data, including scaling and sequence formatting. These choices align with existing literature, which highlights the importance of tailoring preprocessing steps to the characteristics of each model type (Chollet, 2021; Brownlee, 2018).

4.1 Exploratory data analysis (EDA)

In his book, Mount explains that the Exploratory Data Analysis is often an “interviewing” of the data; it is a time for the analyst to get to know it and learn about what interesting things it has to say (Mount, 2021).

To obtain data was analysed to determine its essential traits and connections were investigated following the important principles of EDA, checking data type, structure and correlations (Johnkutty & Davis, 2024).

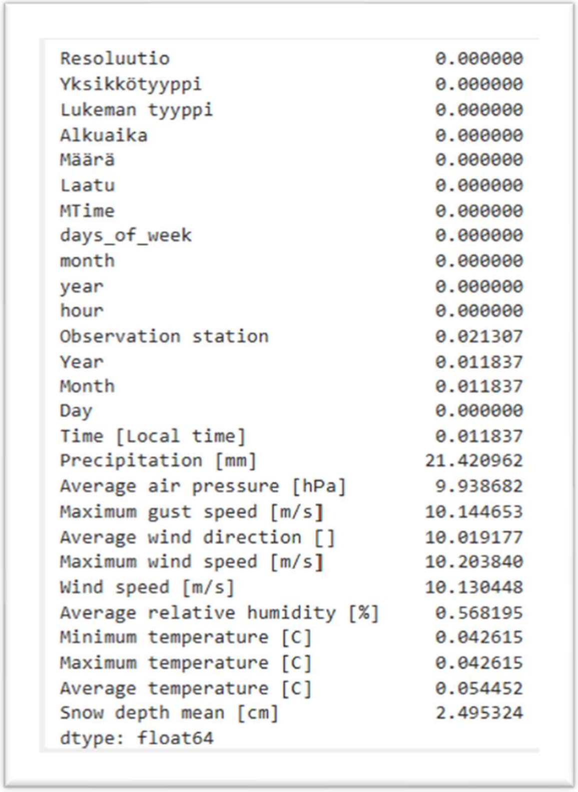
The following steps were performed to gain insights into the data:

- Checking data column types and structure
- Analyzing energy Consumption trends over time
- Analyzing the correlation between temperature and energy consumption

4.2 Missing data and outlier handling

In the dataset, the household energy consumption does not contain missing values. However, the meteorological datasets do have missing values, with the most critical column being “Precipitation [mm]”, which shows 21% missing values in the Jyväskylä dataset. Table.3 Percentage of Missing Values displays the percentage of missing values for each column in the Jyväskylä datasets, where most households are located.

Table 3.Percentage of Missing Values



```
Resoluutio                      0.000000
Yksikkötyyppi                   0.000000
Lukeman tyyppi                  0.000000
Alkuaika                        0.000000
Määrä                           0.000000
Laatu                           0.000000
MTime                           0.000000
days_of_week                    0.000000
month                           0.000000
year                            0.000000
hour                            0.000000
Observation station             0.021307
Year                            0.011837
Month                           0.011837
Day                             0.000000
Time [Local time]               0.011837
Precipitation [mm]             21.420962
Average air pressure [hPa]      9.938682
Maximum gust speed [m/s]       10.144653
Average wind direction []      10.019177
Maximum wind speed [m/s]       10.203840
Wind speed [m/s]               10.130448
Average relative humidity [%]   0.568195
Minimum temperature [C]         0.042615
Maximum temperature [C]         0.042615
Average temperature [C]         0.054452
Snow depth mean [cm]            2.495324
dtype: float64
```

To analyse each of the columns that contained null values in depth, it was necessary to check the values grouped and ordered by day, month and year. This way, it is possible to analyse whether there were time intervals greater than a day without measurements or just a few hours.

Each column was analysed individually by following these steps: First, the null values of each column were separated, copied and isolated in a new dataset. Then, these values were grouped and ordered by day, month and year. The last step was to count the null values based on the day, month and year, so that the last column represented the total null values per day.

The analysis found a high number of null values in the sequences, mainly in the column "precipitation [mm]". Several days without measurements were also found. This missing information in very long sequences, especially in temporal models, can affect the accuracy of the model. Incomplete values can cause problems due to systematic differences between observed and unobserved data. Therefore, it is absolutely important to find the best way to estimate these missing values to ensure high quality of the analysed data (Hawthorne and Elliott, 2005).

A good approach to dealing with missing data is to adopt imputation techniques. These techniques involve replacing missing values with estimates based on the information available in the dataset. Common methods include mean or median imputation, interpolation, regression models, and more advanced approaches such as multiple imputation or machine learning-based techniques. The choice of imputation method depends on the nature of the missing data, the underlying patterns in the dataset, and the specific requirements of the predictive model. Effective imputation not only preserves the temporal continuity of the data, but also minimises potential biases and maintains the statistical integrity of the analysis (Junninen et al., 2004).

For processing meteorological data, a highly recommended approach is to use interpolation based on linear relationships. Linear interpolation estimates missing values using the closest known data points. A commonly used method involves identifying the four closest values and calculating their mean to impute each missing value. Given the significant seasonal variations in meteorological variables, this approach provides a more realistic estimate compared to methods that assume uniformity over time (Soltani and Meinke, 2004).

Mathematically, linear interpolation can be expressed using the following equation

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) \tag{2}$$

This equation above is the Linear Interpolation Equation, where:

- $x_1$and x_2 are the x-values of the know data points,
- y_1 and y_2 are the corresponding y-values,
- x is the x-value where interpolation is applied,
- y is the estimated values at point x.

By applying this formula above, we are estimating the missing meteorological values based on the linear trend between known data points, ensuring smoother transitions in time series datasets.

4.3 Feature engineering and time features

Effective feature engineering is essential in time series forecasting as it helps the model capture temporal patterns and external influences more accurately. In this study, several features were designed to enrich the input data. First, a holiday indicator was introduced to identify national holidays. Such days typically exhibit different consumption patterns due to changes in household routines and occupancy.

Additionally, day-of-the-week features were encoded to capture weekly seasonality, which reflects consistent behavioural differences between weekdays and weekends.

To capture temporal dependencies, a lagged consumption feature was created by shifting the target variable by one hour. Lag features are a common approach in time series modelling, as they allow the model to learn from recent consumption levels (Zhang et al., 2021).

Moreover, two statistical features were computed: the mean and standard deviation of energy consumption over the previous three hours. These features provide local context, helping the model identify short-term trends and volatility in consumption behaviour. By combining calendar-based and temporal features, the dataset becomes more informative and supports the development of a more robust and accurate predictive model.

4.4 Scaling and normalization

Scaling and normalisation are essential preprocessing steps in machine learning, particularly for algorithms sensitive to the range and distribution of input features. In this study, feature scaling was applied to standardise the range of the numerical variables. Specifically, normalisation transforms features to a common scale, typically by centring around the mean and scaling to unit variance, so that no single feature dominates the learning process due to its magnitude (Han, Pei, & Kamber, 2011).

In time series forecasting and other predictive modelling tasks, it is critical to perform scaling based only on the training data. This avoids a problem known as data leakage, which occurs when information from the test set inadvertently influences the model during training. If the scaling parameters, such as the mean and standard deviation, are computed using the entire dataset, including future or unseen test data, the model may gain unfair knowledge of the data it is supposed to predict. This can lead to overly optimistic performance metrics and poor generalisation to real-world data (Kuhn & Johnson, 2013).

To prevent this, scaling parameters were computed solely from the training set and then applied to both the training and test sets. This ensures a realistic evaluation of the model's performance on truly unseen data and upholds the integrity of the validation process.

4.5 Data splitting (train/validation/test)

To effectively develop and evaluate the predictive model, the dataset was initially divided into three subsets: training dataset, validation dataset, and test dataset. The training dataset was used to adjust the model, allowing it to learn patterns and relationships in the data. The validation dataset served as a baseline for tuning hyperparameters and avoiding overfitting during the model development phase. Finally, the test dataset was presented as an unseen dataset, used only to evaluate the performance and generalisation ability of the final model.

Because this is a time series problem, it is crucial to split the data in time order rather than at random. Random splits risk bleeding future information into the past, which can inflate performance metrics. Instead, the first 70 percent of the timeline can serve as training, the next 15 percent as validation and the final 15 percent as test. Alternatively, a simpler 70 percent training and 30 percent test split may be used when validation is handled by rolling-window or expanding-window cross-validation. In each case, the earliest observations train the model, the middle block guides tuning, and the latest block assesses generalisation.

Maintaining this strict chronological separation avoids data leakage and ensures an honest assessment of how the model will perform when forecasting future energy consumption.

To further improve model performance, statistical smoothing techniques were applied using lagged values. Specifically, the mean and standard deviation (std) of the previous 3 hours of energy consumption were calculated and used as additional features. It can better capture short-term variations and patterns in energy use, helping models better adapt to fluctuations in consumption. Statistical smoothing with lagged values not only helps capture fluctuations in consumption but also helps the forecasting model reduce the impact of outliers and noise in the data, capturing seasonal patterns.

By incorporating these lagged statistical values:

- The mean of the previous 3 hours provides insight into recent trends, stabilising sudden fluctuations.
- The standard deviation helps identify volatility in energy consumption, indicating irregular usage patterns.

These statistical features add strength and robustness to the model, as well as reduce noise, thus maintaining useful temporal relationships in the data. Research shows that including lagged statistical features, such as moving average and standard deviation, significantly improves the accuracy of predictions in time series models (Hyndman & Athanasopoulos, 2018).

The Scikit-learn documentation on lagged features also highlights the effectiveness of using moving window statistics to capture trends and seasonality in time series data (Pedregosa et al., 2011).

In the project, Feature-Engine was used because it is very suitable for this stage, as it provides the best practical guidance on implementing lagged statistics in models for time series forecasting. Feature-Engine is a widely known Python library used in these situations, improving the model and its robustness (Feature-Engine, 2023).

To evaluate the individual data of each one of the houses, it was performed the data splitting for each house. A time-based split was performed instead of a random split to preserve the temporal dependencies in the data. The datasets for each house were split into 70% for training and 30% for testing. It was set sequentially to ensure that the past data information was always used to predict future values. This approach is used to avoid data leakage and also ensures that the model generalises better to unseen data.

A common practice in time series data splitting was used in the most recent portion of the dataset as the test set while keeping earlier observations for training. This is used to ensure that the chronological order of the data will remain intact. In that case, it is crucial to keep the chronological order of the data for accurate time-series forecasting (Bergmeir & Benítez, 2012).

Additionally, methods such as expanding window validation or rolling window validation can further evaluate model performance during different periods (Tashman, 2000).

Hyndman & Athanasopoulos' research has shown that maintaining sequential order in time series forecasting significantly improves model reliability and reduces errors associated with data leakage (Hyndman & Athanasopoulos, 2018). Also, Scikit-learn's guidelines on time series cross-validation emphasise the importance of avoiding a random split to preserve patterns in temporal data (Pedregosa et al., 2011).

4.6 Data merging

Before merging the dataset, the columns, which contained date and hour information were first converted from text to a proper date-time format, as these values were initially recognised as text. The weather data was then checked for duplicate values, and when duplicates were found, the row containing the most complete information was kept while the others were removed. After handling duplicates, the energy consumption data was merged with the corresponding weather data from each dataset.

Since it was not possible to collect all the weather data in a single file, the merge process and duplicate check had to be repeated multiple times. Additionally, a function was created to verify whether a real value was available for replacement. If a valid value was found, it was used; otherwise, the original NaN values remained unchanged.

The initial step of Exploratory Data Analysis always involves examining the data types. Checking the types and structure of the data columns, as well as checking the integrity and consistency of the data to ensure that all variables have been formatted correctly for further analysis. The dataset consists of:

- Time-based attributes (e.g., timestamp, year, month, and hour of the day).
- Energy consumption values are measured in kilowatt-hours (kWh).

- Meteorological variables such as temperature, humidity, wind speed, and precipitation.

A summary of the dataset confirmed that the data structure and type were correct. Energy consumption was correctly stored as continuous numeric values, while time-related variables were correctly formatted as date and time objects. The presence of missing values in some timestamps required interpolation techniques to ensure a continuous time series for accurate forecasting. Additionally, outliers in energy consumption were examined to differentiate between genuine periods of high demand (e.g., extreme cold weather) and potential data recording errors.

4.7 Data preprocessing in machine learning: adapting techniques to suit different model architectures.

When developing predictive models for time series data, it is crucial to adapt the preprocessing steps to the specific architecture of each model and respect them. This project considered and evaluated the necessary preprocessing adaptations specific to each model architecture. For example, Recurrent Neural Networks (RNNs) such as Long Short-Term Memory (LSTM) networks and Transformer models have distinct data requirements compared to tree-based models such as LightGBM (LGBM).

LSTM and Transformer models are designed to handle sequential data, making them well-suited for time series forecasting.

The encoding method significantly impacts algorithm performance for cyclical data in machine learning, such as months and hour-of-day (Mahajan, Singh & Bruns, n.d.).

For months and hours, sine and cosine transformations were applied to capture cyclical patterns in the data, allowing the model to understand the periodicity of energy usage. For example, January is closer to December than January is to June in terms of energy usage. Since December and January are winter, while June and July are summer.

Scaling was applied to the numerical values so that the values would fall within a specific range. Standardisation (Z-score normalisation) was selected at this point in the study. Z-score was selected as the preferred scaling method over Min-Max scaling. This decision was based on the nature of the dataset, which includes environmental variables such as temperature and humidity. Since the available data spans only three years, it is possible that the observed minimum and maximum values do not represent the true extremes of these features. According to Han et al., this can negatively impact model performance by introducing unexpected numerical representations that the model was not trained on. Consequently, using Min-Max scaling may lead to issues where future test data contains values outside the originally observed range, potentially resulting in values beyond the expected range [0.1][0.1][0.1] (Han et al., 2011).

Mathematically, Min-Max scaling is defined as:

(3) $$X' = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

The equation above is the Min-Max scaling Formula, where:

- X is the original feature value,
- $Xmin$ and $Xmax$ are the minimum and maximum values of that feature in the training set,
- x' is the scaled value between 0 and 1.

However, if a new data point exceeds $Xmax$ or falls below $Xmin$, the transformation will produce values greater than 1 or less than 0, respectively, which can lead to unexpected model behaviour.

To mitigate this issue, standardisation (Z-score normalisation) was chosen. Standardisation transforms data such that the mean becomes 0 and the standard deviation becomes 1, making it more robust to unseen extreme values (Bishop, 2006). The formula for standardisation is:

(4) $$Z = \frac{X - \mu}{\sigma}$$

The above formula is Standardization, where:

- X is the original feature value, the data point
- μ is the mean of the feature in the training of the dataset,
- σ is the standard deviation of the feature in the training dataset,
- Z is the standardised value.

According to Gareth James, standardisation ensures that the transformed feature distributions remain consistent even if new data points exceed the previous training range, differing from Min-

Max scaling. This characteristic is mostly beneficial in forecasting applications where environmental variables can fluctuate due to seasonal variations or climate changes (James et al., 2013).

It is vital to maintain consistency in feature scaling over different datasets. At this point in the study, standardisation parameters (mean and standard deviation) were computed using only the training data. These parameters were then applied to the model to transform the training and subsequent datasets, thus following best practices for the generalisation of machine learning models. This approach was used to ensure that unseen future data is processed under the same distribution as the training set, improving model reliability. (Goodfellow et al., 2016). At that time, the fit_transform( ) function was applied exclusively to the training data, while the transform( ) function was mainly used for validation and test datasets to avoid data leakage (Geron, 2019).

A function, create_sequence, generates a sequence of data for a given 24 hours for each feature value over the previous 24 hours, while the target values are for the current hours.

LightGBM is a gradient boosting framework that utilises tree-based learning algorithms. Unlike LSTM and Transformer models, LightGBM does not inherently account for temporal order, requiring specific pre-processing steps.

It was time to encode months and hours as categorical variables. At that point, temporal features like months and hours were incorporated into the machine learning models, mainly tree-based models such as LightGBM, which require careful encoding to capture patterns inherent in time series data effectively. One common approach is one-hot encoding, which transforms categorical variables into binary vectors, allowing the model to interpret each category distinctly without assuming any ordinal relationship (Brownlee, 2016).

One-hot encoding increases the number of features, which can lead to higher memory usage and computational complexity; a proper encoding of cyclical features ensures that the model does not assume a false ordinal relationship, thereby improving the accuracy of predictions (Skforecast, 2023).

Weekdays were added as categorical variables to capture weekly consumption patterns. This reflects routines for each day of the week. Without beliefs is ordinal data.

The previous hour's energy consumption was added to the dataset as a feature. This technique of including lagged variables improves the prediction because it can capture temporal trends. This can increase the accuracy of the model, as recurring peaks and troughs in consumption are likely to be more influenced by the daily consumption patterns of the household and individuals. Incorporating lagged values as features is particularly effective in energy forecasting, as demonstrated in research where lagged data improved the accuracy in forecasting short-term fluctuations and demand (Lu, Yuan, Schwartz, & Benjamin, 2007).

4.8 Final dataset

The final dataset brings together temporal, historical and environmental information for each energy reading. Temporal context is captured by the month, the weekday, the hour and the year. Short-term usage history comes from lagged consumption values, three moving-average windows built on those lags and the corresponding three-hour standard deviations. Each record also carries a unique house identifier to track individual household patterns. Environmental context is provided by the daily average temperature, daily average wind speed, daily average humidity, daily snow depth in centimetres and daily cloud cover. This combination of features gives the model a detailed understanding of when, how and under what conditions households use energy.

4.8.1 Analyzing energy consumption trends over time

To better understand energy consumption patterns, total monthly energy consumption was analysed for each of the analysed homes over a five-year period (2019–2024). Figure 3 below ("Monthly Energy Consumption by Home") illustrates energy usage trends for each of the individual homes.

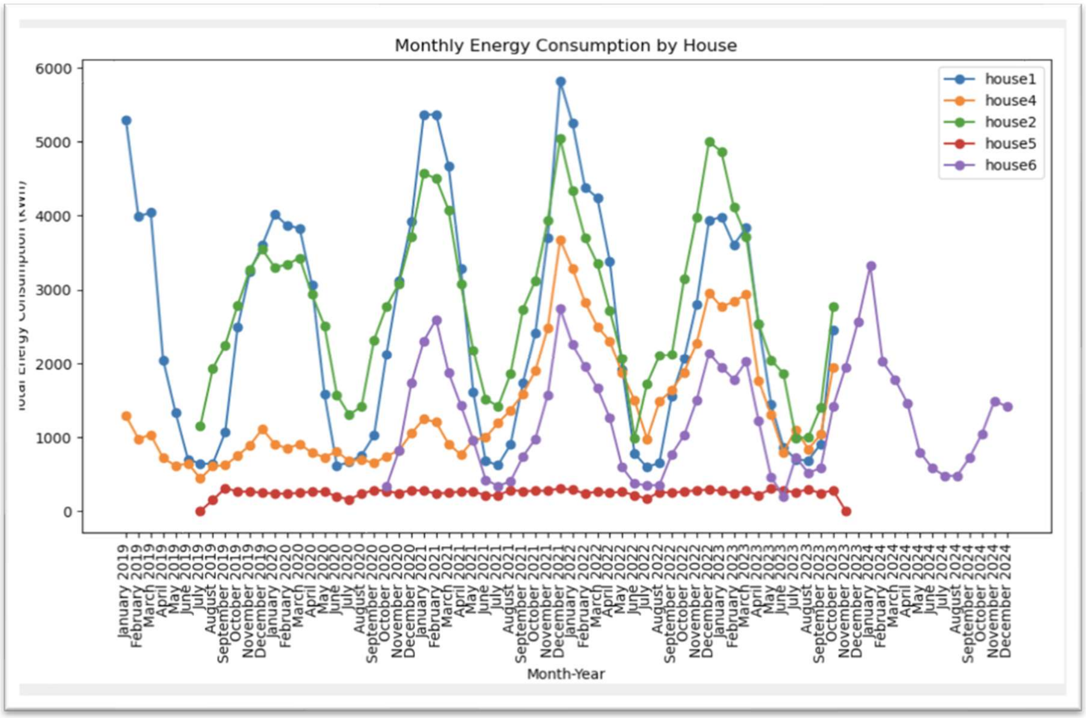


Figure 3 Monthly Energy Consumption by House

Key Observation from Figure 3:

- Seasonality: Energy consumption exhibits seasonal variations, where it clearly shows a higher energy demand in the winter months and the lowest energy consumption in the summer months. This pattern demonstrates that heating systems significantly influence household energy use as expected, mostly in the colder seasons (Swan & Ugursal, 2009).

- Long-Term Trends: Some houses exhibit gradual increases in energy consumption over time, which may be due to changes in occupancy, appliance usage, or variations in energy efficiency.

- Differences Between Houses:
  - House1, House2, and House6 display higher energy consumption and greater fluctuations, likely due to larger household size, poor insulation, or higher heating needs.

- House4 shows moderate energy usage, lower and most stable consumption levels over the years, suggesting either higher energy efficiency, reduced heating dependency.

- While House5 has the lowest and most stable consumption levels during the years, because it is a cottage, House5 uses fewer electronic devices, has lower occupancy, because it is mainly used in the summertime.

The confirmation of the strong seasonal patterns observed implies the need to incorporate these seasonal characteristics into the predictive modelling process to ensure that the model can effectively capture fluctuations in energy demand based on the time of year (Hyndman & Athanasopoulos, 2021).

4.8.2 Correlation analysis between temperature and energy consumption

To better understand the relationship between outdoor weather conditions and household energy use, a Pearson correlation analysis was conducted between temperature and total energy consumption for each household. Figure 4 below presents the computed correlation coefficients, which quantify the degree of linear dependence between these two variables. Thus, the correlation coefficient ranges from -1 to 1, where values closer to -1 indicate a strong inverse relationship, meaning that the more the temperature decreases, the more energy consumption increases (Shumway & Stoffer, 2017).

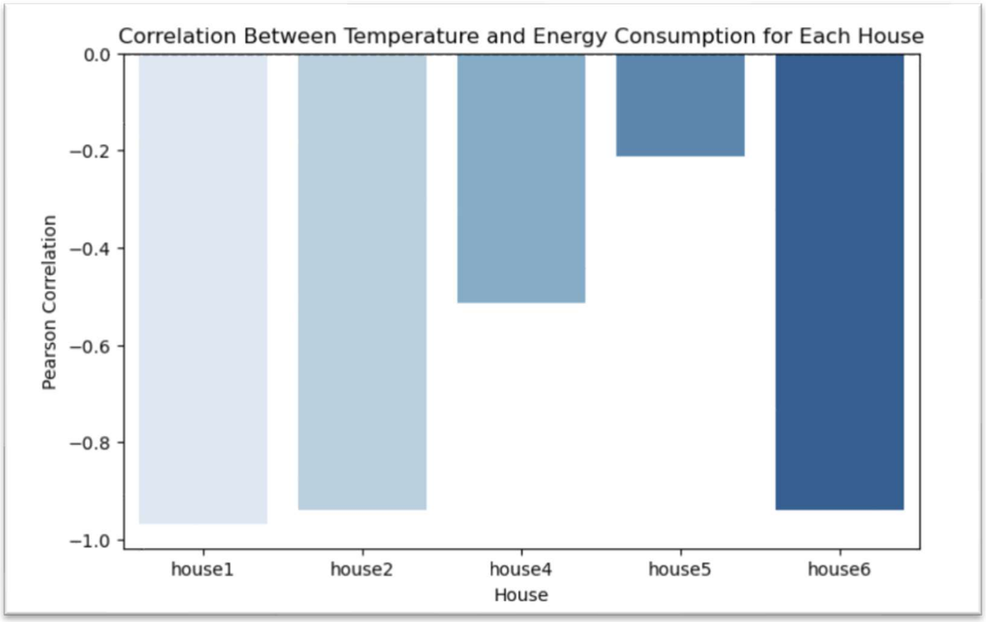


Figure 4. Correlation between temperature and energy consumption for each house

The results of the correlation analysis between temperature and energy consumption indicate a consistent negative correlation among all houses, although the strength of this relationship varies between houses. House1, House2, and House6 exhibit the strongest negative correlations, suggesting that energy consumption in these houses is highly sensitive to temperature fluctuations. This is likely due to a greater reliance on heating systems during the colder months, leading to higher energy consumption in winter and lower energy consumption in warmer months (Aghabozorgi et al., 2015; Hyndman & Athanasopoulos, 2021).

On the other hand, still analyzing the image, we can see that House4 presents a moderate negative correlation, indicating that, although temperature still plays a role in energy consumption, other factors such as household behaviour, high thermal insulation systems, smart appliances and alternative heating systems can also influence energy demand. In contrast, observing House5, we can see that the correlation is considerably weaker, suggesting that changes in external temperature have a minimal impact on its energy use. This can be attributed to several factors, such as lower occupancy since it is a “mökki” (cottage house) used primarily in the summer, and with a smaller number of appliances and electronic devices (Swan & Ugursal, 2009).

The findings from this correlation analysis provide critical insights into the predictive modelling process:

- The strong negative correlation in most houses reinforces the importance of including temperature as a primary feature in the energy forecasting model.
- The variation in correlation strength across different houses suggests that a generalised model may not perform equally well for all households. Instead, house-specific modelling approaches or the inclusion of additional features (e.g., occupancy patterns, appliance usage, and insulation quality) may be necessary for improving prediction accuracy.
- The weak correlation observed in House 5 implies that other explanatory variables, such as household schedules, solar panel usage, or smart home energy optimisations, could be influencing its energy demand more than temperature alone.

By incorporating these insights, the forecasting model can be better structured to capture temperature-dependent fluctuations in energy consumption, ensuring higher predictive accuracy and improved adaptability to different household energy profiles.

4.8.3 Time-series decomposition of the energy consumption

A time series decomposition was performed to analyse the underlying patterns of the energy consumption data. As can be seen in Figure 5 below, four elements had the data decomposed: Original Energy Consumption, Consumption Trend, Trend Component, and Residual Component. They were represented in four different colours: Original Energy Consumption in blue, Trend Component in green, Seasonal Component in red, and Residual Component in purple. This decomposition process is responsible for separating the Original Energy Consumption Data into three main energy consumption components: trend, seasonality, and residual variance. This separation allows a better understanding of the energy use patterns and, therefore, improves the forecast accuracy.

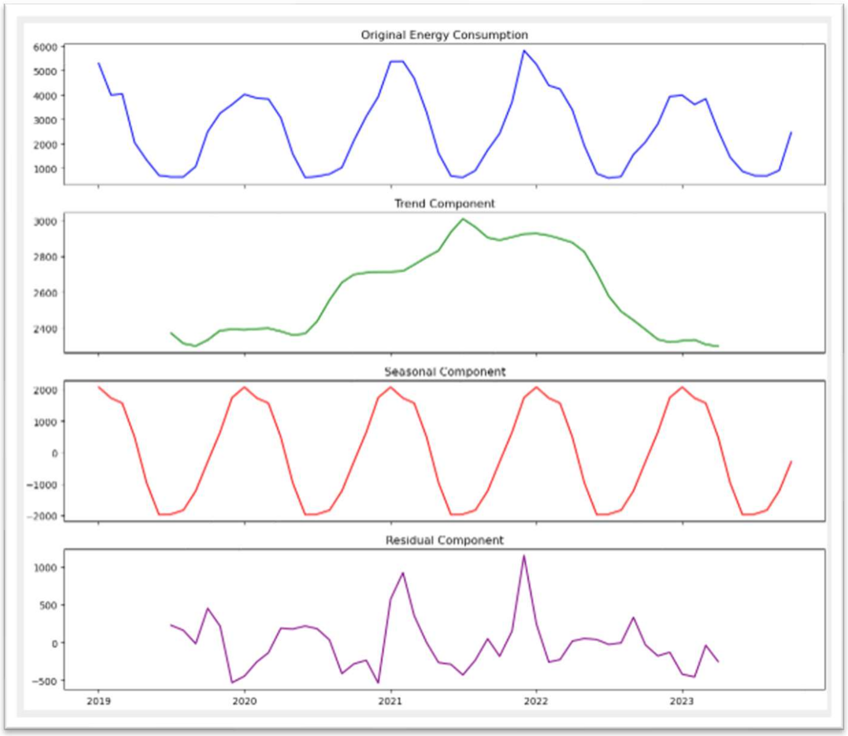


Figure 5. The graph of the Times-Series Decomposition of the Energy Consumption

The original energy consumption, as can be seen in blue in the first graph of Figure 5, represents the raw energy consumption data that has been extracted over time. Observing the graph, the fluctuations indicate that the seasonal pattern plays a very strong influence on the energy consumption. The heavy peaks and the troughs in the graph are periodic areas related to the oscillation of the energy consumption. It suggests that it occurs due to external factors like seasonal habits, changes in temperature, or daylight hours.

The consumption trend component, shown in the second image graph in green, reveals the long-term direction of energy consumption. This data suggests an upward trend in specific periods, most likely due to long, harsh winters with very low temperatures or even an increase in energy demand. The declining sections may indicate improvements in energy efficiency or even a reduction in energy consumption during milder seasons. This data is important for trying to understand this oscillation in consumption trends so that long-term energy production and distribution planning can be truly efficient.

In the third graph of Figure 5 in red, we can see the data of the Seasonal Component isolated. The graph shows a big fluctuation in energy consumption during the different seasons. This graph highlights the repeating patterns occurring at fixed intervals. The pronounced oscillations confirm that energy consumption follows a highly seasonal structure. There are large peaks during the winter months and significant troughs where we have the summer months. The energy consumption makes the transition during the intermediate seasons, like Autumn and Spring. To improve forecasting models and ensure the accuracy of the model, it is essential to identify these patterns that capture the expected fluctuations.

The final Residual Component plot in purple in Figure 6 illustrates components that are not yet explained by the seasonal patterns nor the consumption trends. However, these data account for irregular variations that have to be investigated. The presence of unexpected fluctuations plotted in the graphic shows the noise and spikes that can be the result of extreme weather events, changes in energy taxation policies, or even anomalies in the data record. The analysis of these residuals could help to detect the discrepancies in consumption data in an attempt to refine forecasting models.

In conclusion, Figure 5 visually demonstrates the importance of analyzing energy consumption data in isolation. This allows us to understand the influence of seasonal climate factors on energy consumption. The data separated by time series decomposition could capture important aspects of the dataset, like long-term trends, seasonal components, and unexpected consumption fluctuations. It is essential to consider all of these factors that influence the results individually and their combination to develop an accurate and robust energy consumption forecasting model that, in the future, can assist in better energy management strategies.

5 Model development

Here will be presented the models of electricity consumption prediction will be presented. The first model is the Architecture and Attention Mechanism for Energy Consumption Forecasting Model, the second one is the Long Short-Term Memory Model (LSTM) for Energy Consumption Forecasting, and the third model is the Light Gradient Boosting Machine (LightGBM).

5.1 Model 1: attention and Long Short-Term Memory-based Model

The first prediction model created was the AttentionLSTM model, which combines a Long Short-Term Memory layer with a head-attention mechanism. Energy consumption forecasting is a crucial aspect of efficient energy management, enabling power plants and grid operators to optimise resources, reduce costs, and enhance sustainability. In this thesis, a Long Short-Term Memory (LSTM) model with an attention mechanism is utilised to predict hourly energy consumption based on historical data of houses and meteorological factors. This section details the architecture of the model and explains how the attention mechanism enhances its predictive performance.

The AttentionLSTM model architecture was based on a deep learning framework that combines multiple layers. The input layer, Long Short-Term Memory (LSTM) layer with an attention mechanism layer, dense fully connected layer, and finally the output layer was used to effectively capture temporal dependencies and highlight key features in the dataset.

The AttentionLSTM model architecture has the following layers:

1. Input Layer: The model receives time-series data, including past energy consumption values, temperature, wind speed, air pressure, and categorical variables such as working days or holidays and hour-of-day indicators.
2. LSTM Layers: These layers were responsible for processing sequential data, capturing long-term dependencies and patterns in energy consumption trends.
3. Attention Layer: The attention mechanism dynamically assigns weights to different time steps, allowing the model to focus on the most relevant past observations.
4. Fully Connected (Dense) Layer: This layer integrates the context vector from the attention layer with the LSTM output to refine predictions.
5. Output Layer: The final prediction is obtained, representing the estimated energy consumption for the next time step.

This diagram below in Figure 6 illustrates the model’s structure, including input features, LSTM layers, an attention mechanism, and the final prediction layer. This diagram illustrates the flow steps of data through the different layers:

1. **Input Layer** (light blue): Accepts the time-series data.

2. **LSTM Layer** (light green): Processes sequential dependencies in the data.
3. **Attention Layer** (yellow): Dynamically assigns importance to relevant past time steps.
4. **Dense Layer** (orange): Combines the attention-enhanced context with the LSTM output.
5. **Output Layer** (red): Generates the final energy consumption forecast.

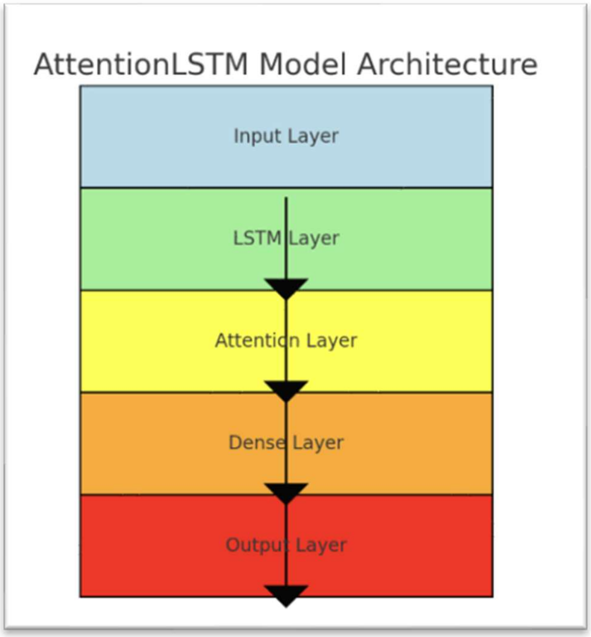


Figure 6. Architecture of the AttentionLSTM Model

The attention mechanism is integrated into the LSTM model to overcome its limitation of treating all past time steps equally. Instead of relying on the final hidden state of the LSTM, attention allows the model to dynamically assign importance to different past observations. The process is as follows:

1. Score Calculation
2. Softmax Normalisation
3. Context Vector Computation

4. Final Prediction

Score Calculation: Each hidden state in the LSTM sequence is assigned a relevance score based on its impact on the target prediction. The score function is computed as:

(5)
$$\text{score}_t = V^T \tanh(W_1 h_t + W_2 h_s)$$

The equation above is the Score Calculation Computational where:

- ht: represents the current hidden state.
- hs: represents past hidden states,
- W_1 and W_2: are learnable weight matrices.
- V^T: a transposed vector used to project the hidden layer to a scalar score
- the hyperbolic tangent activation function, introducing non-linearity

3. Softmax Normalisation: Converts scores into attention weights:

(6)
$$\alpha_t = \frac{e^{\text{score}_t}}{\sum_s e^{\text{score}_s}}$$

The equation above is the Softmax Normalisation Computational Equation, where:

- αt: is the attention weight for timestep t.
- e_{scoret}: Exponentiate the score of t to ensure it's positive.
- $\sum s\, e_{scoret}$: Sum up the exponentials of all scores over the index s

4. Context Vector Computation: A weighted sum of past hidden states is calculated to form the context vector computational equation:

(7)
$$c_t = \sum_s \alpha_s h_s$$

The Context Vector Computational is a weighted sum of past hidden states, used to refine the prediction, where:

- C_t: is called the context vector at time t.
- αs: is the attention weight for each position s

- h_s: is the hidden state at position s.

5. Final Prediction: The model concatenates the context vector with the LSTM output and passes it through a dense layer to generate the final forecast:

$$y_t = \mathrm{Dense}\big(\mathrm{concat}(c_t, h_t)\big) \tag{8}$$

The Dense Layer Computational Equation above combines the attention-enhanced context vector and the current hidden state, where:

- c_t*:* is the context vector,
- h_t : is the current hidden state at time t,
- $\mathrm{concat}(c_t, h_t)$: means concatenating the two vectors,
- *Dense* $\big(\mathrm{concat}(c_t, h_t)\big)$: basically, a simple neural network layer that transforms the input.

There are many studies providing evidence for the effectiveness of integrating attention mechanisms with LSTM models for electricity prediction:

- Electricity Consumption Prediction Using LSTM with Attention: This study showed a 6.5% increase in prediction accuracy when an attention mechanism was added to an LSTM model. The attention mechanism helped the model focus on the most influential input sequences, improving forecasting performance (ResearchGate, n.d.).
- Hybrid Attention-Enhanced Deep Learning for Hourly Energy Forecasting: The integration of attention mechanisms in LSTM-based forecasting models resulted in superior predictive performance, achieving an R^2 score of 0.992, demonstrating high accuracy (Sciences Force, 2024).
- ANN-LSTM-A Model for Water Consumption Prediction: Although focused on water consumption, this research highlights how combining an ANN with an LSTM model and an attention mechanism significantly improved long-term dependency handling and prediction accuracy (MDPI, n.d.).

- Zhang et al. (2020) demonstrated that the addition of an attention mechanism improved energy consumption forecasting by allowing the model to focus on critical input time steps, reducing forecasting errors and improving interpretability.

These findings validate the use of attention-based LSTM models for forecasting time-series energy consumption and justify the inclusion of such mechanisms in this research.

Integrating attention mechanisms into LSTM models significantly enhances energy consumption forecasting by dynamically weighting past data points, improving accuracy, and capturing temporal dependencies. Supporting literature confirms the effectiveness of this approach, highlighting its potential for real-world energy management applications.

5.2 Model 2: Transformer-based Model for energy consumption forecasting

While the first model integrates an attention mechanism with LSTM layers, an alternative approach involves using a stacked LSTM model with multiple hidden layers to capture temporal dependencies and improve forecast accuracy. This model follows a sequential deep learning architecture, leveraging LSTM layers, dropout regularisation, and fully connected layers.

The architecture of the model is structured as follows:

1. LSTM Layer (128 units): Captures temporal dependencies in the sequential energy consumption data. It processes input sequences and extracts long-term patterns fundamental for forecasting.

2. Dropout Layer (50%): Regularisation technique that randomly deactivated 50% of neurons during training to prevent overfitting and improve generalisation.

3. Dense Layer (ReLU activation): Fully connected layer reducing dimensionality while maintaining critical features extracted from LSTM outputs.

4. Dropout Layer (20%): Additional regularisation to prevent model overfitting by randomly deactivating neurons during training.

5. Dense Layer (ReLU activation): Further dimensionality reduction while extracting meaningful relationships in the transformed data.

6. Dense Output Layer (Linear activation): Predicts a single continuous value, representing the estimated energy consumption for the next time step.

The model was compiled using the Adam optimiser and the Mean Squared Error (MSE) loss function, ensuring efficient weight updates while minimising prediction errors. Training was performed using a batch size of 64 and 10 epochs, with validation data used to monitor generalisation performance.

The architecture of the trained LSTM model is visually represented in Figure 7 below, illustrating the sequential flow of data through different layers. Each layer plays a specific role in refining the input data and improving forecasting accuracy.

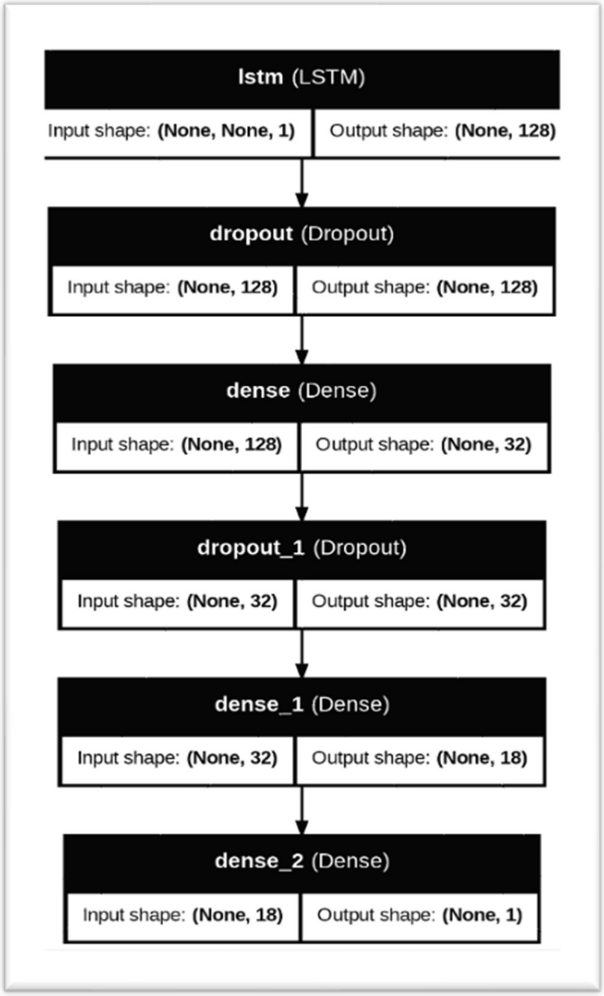


Figure 7 . LSTM Model Architecture Diagram

Several studies provide evidence for the effectiveness of LSTM models in energy forecasting:

- LSTM-Based Time-Series Forecasting for Energy Consumption: Research has demonstrated that LSTM models outperform traditional statistical models, such as ARIMA, in predicting energy consumption due to their ability to capture long-term dependencies (Siami-Namini et al., 2019).
- Deep Learning for Energy Demand Prediction: Studies have shown that deep learning models, including LSTM networks, achieve lower forecasting errors compared to conventional regression models, mostly when integrating meteorological variables (Yan et al., 2020).
- The Impact of Dropout in LSTM Models: Empirical results highlight the importance of dropout layers in preventing overfitting and improving generalisation in time-series forecasting tasks, particularly in energy demand prediction (Zhang & Yan, 2021).
- Performance Evaluation of LSTM for Energy Forecasting: Comparative analyses indicate that LSTM models, when combined with optimized hyperparameters and batch normalisation, can enhance predictive accuracy and model robustness (Gers et al., 2017).

These findings validate the choice of LSTM-based models for energy consumption forecasting and support their integration in this research.

This structured approach ensures that the model effectively captures short- and long-term dependencies in energy consumption data, leading to more accurate predictions. The evaluation results of this model and a comparison with other approaches will be discussed in the following sections.

The implemented LSTM model demonstrates strong potential for accurately forecasting residential energy consumption by effectively capturing long-term dependencies in time-series data. The integration of dropout layers enhances model robustness by preventing overfitting, while the structured use of dense layers refines feature extraction. Empirical studies support the effectiveness of LSTM architectures for energy forecasting, reinforcing the decision to utilise this approach.

Despite its advantages, the model may benefit from hyperparameter optimisation, additional feature engineering, and extended training time to further improve performance. Future research could explore hybrid models, such as combining LSTM with attention mechanisms or tree-based models, to enhance forecasting accuracy and interpretability. Overall, the LSTM-based approach provides a promising solution for improving energy consumption prediction and optimising resource management.

To complement the deep learning approach, a Light Gradient Boosting Machine (LightGBM) model was trained to predict hourly energy consumption. LightGBM is a tree-based ensemble learning method that has demonstrated high efficiency and predictive accuracy in time-series forecasting. This section details the architecture of the model and explains how each hyperparameter contributes to improving predictive performance.

5.3 Model 3: LightGBM Model architecture

The LightGBM model was configured with the following hyperparameters in this project:

1. Metric (RMSE): Root Mean Squared Error (RMSE) was chosen as the evaluation metric, as it effectively penalises large deviations in energy consumption predictions.
2. Number of Estimators: The number of boosting iterations, determining the depth of the model's learning process.
3. L2 Regularisation: Helps prevent overfitting by adding a penalty term to the loss function.
4. L1 Regularisation: Controls model complexity by encouraging sparsity in feature importance.
5. Column Subsampling: Uses a fraction of available features for training, reducing overfitting.
6. Row Subsampling: Selects a subset of training data per iteration to improve generalisation.
7. Learning Rate: Controls the step size in weight updates to balance convergence speed and accuracy.
8. Maximum Depth Limits tree depth to prevent excessive complexity and overfitting.
9. Number of Leaves: Controls the number of terminal nodes in each decision tree.

10. Minimum Child Samples: Defines the minimum number of observations required to split a node.
11. Maximum Bin: Sets the number of discrete bins for continuous features to enhance computational efficiency.
12. Categorical Feature Regularisation: Applies additional L2 regularisation to categorical variables.

The model was trained using the LGBMRegressor function from LightGBM and fitted on the training dataset using 1100 boosting iterations. The validation set was used to monitor model performance, and predictions were made on the test set.

The architecture of the trained LightGBM model is visually represented in Figure 8 below, illustrating the hierarchical structure of decision trees and their contribution to energy consumption forecasting.

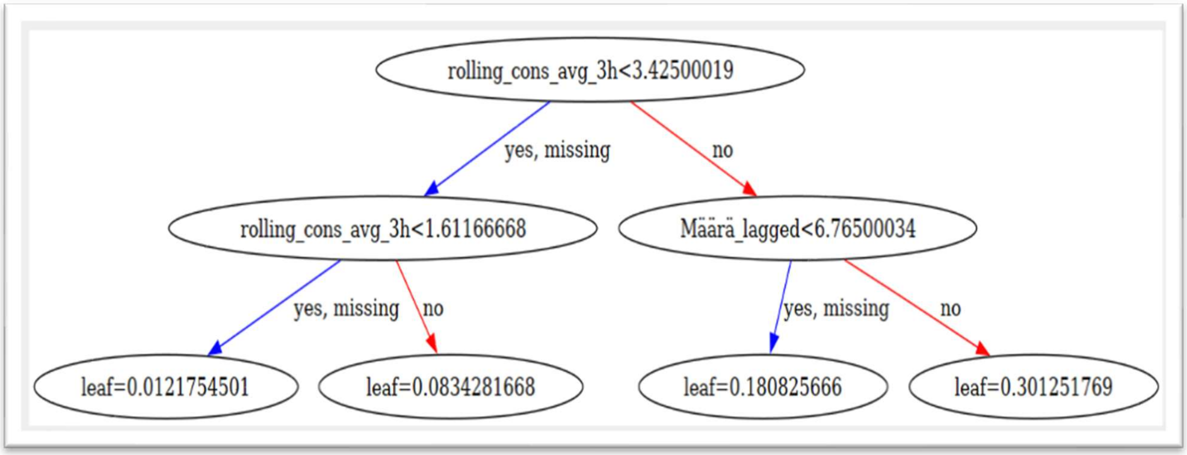


Figure 8. LightGBM Model Architecture, hierarchical flow visualization

Many studies provide evidence of the effectiveness of LightGBM Model in energy prediction:

- LightGBM has been found to outperform traditional regression models and deep learning techniques due to its efficiency in handling large datasets and complex feature interactions.

The gradient boosting LightGBM models help weak learning models become stronger (Ke et al., 2017).

- Empirical studies indicate that models using LightGBM achieve lower prediction errors compared to neural networks when applied to structured time series data. Tree-based models for time series forecasting, such as LightGBM, have lower errors (Shwartz-Ziv & Armon, 2022).
- The research conducted in the paper "Energy Consumption Forecasting Based on explainable AI-Enabled LightGBM" highlights the importance of tuning hyperparameters to improve predictive accuracy and model interpretability. Optimising hyperparameters such as learning rate, tree depth, and feature selection is important to improve the model (Munir et al., 2023).
- Comparative analysis of the findings related to LightGBM and LSTM models in energy forecasting suggests that LSTM models capture temporal dependencies, but LightGBM can outperform deep learning models when feature engineering is optimized (Florek & Zagdański, 2023).

All these studies reinforce the superiority of using LightGBM to predict energy consumption, validating its integration in this research.

6 Model evaluation

After the work is done, the three models were tested, and it is really important to evaluate them. It is crucial to evaluate their performance to verify their accuracy in predicting electricity consumption. This evaluation and comparison are to ensure that the predictions of the models are reliable and applicable to new data.

This is the part where the performances of all three models will be deeply analysed using metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). It is also the part to compare the Attention-based LSTM, Attention-free LSTM, and LightGBM models, highlighting their strengths and limitations. The goal is to identify the most efficient model for energy prediction consumption, balancing accuracy, scalability, and computational cost.

6.1 Evaluation strategy

Evaluating the performance of forecasting models is a critical step in assessing their reliability and effectiveness in real-world applications. In this study, three different machine learning models were implemented to predict residential energy consumption. Each model leverages different mechanisms to capture patterns in the dataset, and their comparative evaluation provides insights into their respective strengths and weaknesses.

- Long Short-Term Memory (LSTM) with Attention,
- LSTM without Attention,
- and Light Gradient Boosting Machine (LightGBM)

The discussion in this section aims to analyse the effectiveness of these models based on established performance metrics and to interpret their results in the context of energy consumption forecasting. By comparing their computational efficiency, accuracy, and scalability, we can identify the most suitable approach for future applications.

This section analyses the effectiveness of the models based on performance metrics, stability and interpretation of the results for forecasting energy consumption. This was done by comparing computational efficiency, accuracy, and scalability in order to identify the best approach for future applications.

6.2 Model results, comparative analysis and the forecasts visualisation

The three models LSTM with Attention, LSTM without Attention, and LightGBM. Were evaluated based on four main factors:

1 Prediction Accuracy – Measured using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R^2), indicating the deviation of predictions from actual energy consumption values.

2 Computational Efficiency – Evaluates training time and model complexity, assessing the feasibility of real-time energy forecasting.

3 Feature Importance and Interpretability – Examines the ability of each model to identify significant energy consumption patterns and provide interpretable results for decision-making.

4 Scalability – Analyse how well each model performs when applied to larger datasets or different regions.

The evaluation results demonstrated that the LightGBM model consistently outperformed both LSTM-based models, it was illustrated in the graphic below. Specifically, LightGBM achieved the lowest RMSE (0.407), MAE (0.251) and the highest R² score (0.969) on the test dataset. This indicates a high level of predictive accuracy and generalisation to unseen data. Additionally, LightGBM offers faster training times and easier interpretability through feature importance plots, making it particularly suitable for large-scale or real-time applications.

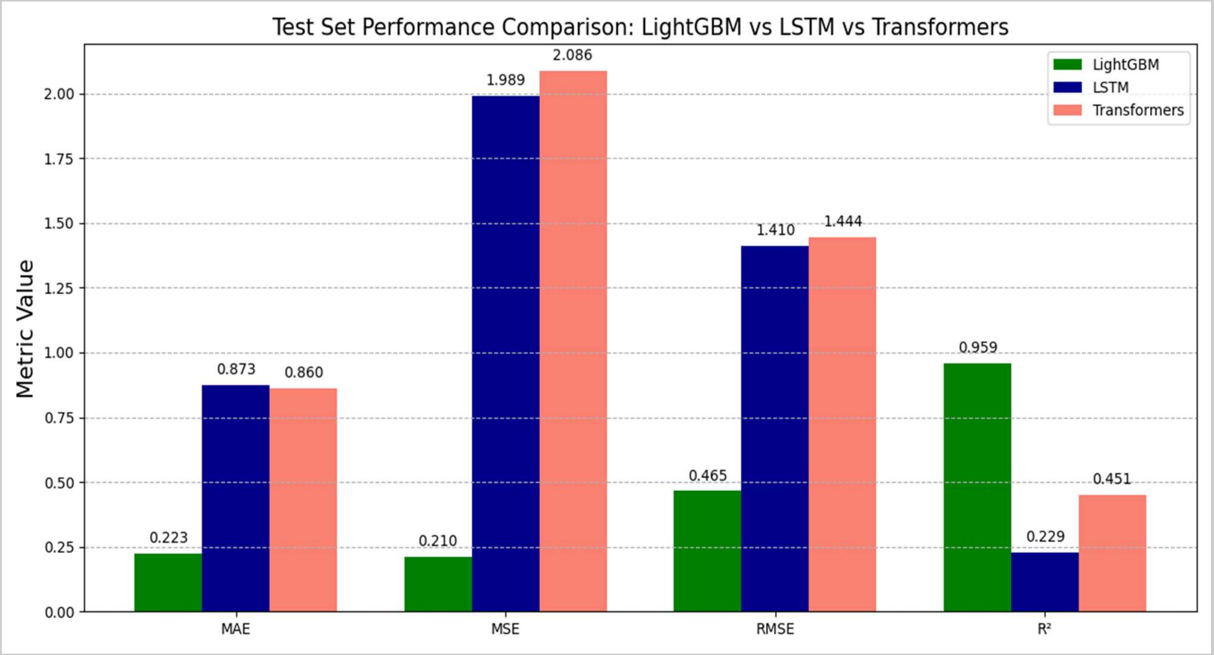


Figure 9. Comparison of test set performance across the models

The graphic above (Figure 9) is a Comparison of test set performance with LightGBM model (green), LSTM model (orange), and Transformer models (purple) based on MAE, MSE, RMSE, and

R^2 metrics. Lower values of MAE, MSE, and RMSE and higher R^2 indicate better performance. LightGBM outperforms the other models over all metrics.

As illustrated in Figure 9, LightGBM (in green) delivers the most reliable performance across all evaluation metrics. Both LSTM-based models, mostly the version without Attention, exhibited higher error values and lower predictive power. While the Transformer model showed some improvement over standard Long Short-Term Memory (LSTM) in terms of R^2 (0.451 vs. 0.229), its performance was still way lower than that of LightGBM. These findings strongly support the use of LightGBM for energy consumption forecasting in structured, tabular datasets.

Different studies were able to provide theoretical evidence to prove the effectiveness of integrating attention mechanisms into LSTM models and the superiority of the Light Gradient Boosting Machine techniques in tabular datasets:

- Attention Mechanisms in LSTM: According to research, incorporating attention mechanisms into LSTMs has shown to improve time series forecasting by dynamically weighting important time steps. (Vaswani et al., 2017).
- LightGBM for Time-Series Forecasting: Studies indicate that tree-based models often perform better than deep learning models well on structured datasets. This advantage is largely due to their efficient handling of categorical variables (Ke et al., 2017).
- Comparative Analysis of Boosting Models vs Deep Learning: The findings suggest that although LSTM models are able to capture temporal dependencies effectively, LightGBM remains strong performer when feature engineering is properly optimized (Florek & Zagdański, 2023).

6.3 Discussion of strengths and weaknesses

The model performance evaluation revealed that LightGBM had the better performance among all models tested for the energy consumption prediction task. LightGBM achieved the best accuracy, required less training time, and produced more interpretable outputs compared to LSTM-based models. While LSTM with Attention showed improvement over standard Long Short-Term Memory (LSTM). However, both models were less accurate and more computationally intensive.

For practical applications, the selection of a model depends on the specific requirements:

- If maximum accuracy and interpretability are desired, LightGBM is the most suitable.
- If sequence modelling or more dynamic forecasting is needed in future research, LSTM with Attention can be explored further, particularly with larger or more complex temporal datasets.

Future work can investigate hybrid approaches, combining the interpretability and speed of LightGBM for feature selection with the sequence modelling capability of Long Short-Term Memory (LSTM) networks to achieve a balanced solution.

7 Discussion

7.1 Interpretation of the findings

The implemented LightGBM model demonstrated strong performance in predicting energy consumption. The LightGBM model is capable of handling high-dimensional features and non-linear relationships, making it a suitable alternative to deep learning models. Compared to LSTM, LightGBM provided better training time values, the selection proved to be more robust, and with better features and data interpretation, the LightGBM model maintained high predictive accuracy.

Despite these advantages, it would be interesting to explore hybrid models that combine LightGBM with deep learning architectures Long Short-Term Memory or even combine it with attention-based transformers in the future. This could highlight sequential dependencies and the importance of structured features. In addition, new hyperparameter adjustments could refine the model's performance. Overall, LightGBM proved to be a great approach because it is scalable and efficient for energy forecasting, capable of contributing to more accurate and data-driven energy management strategies.

7.2 Implications for real-world energy management

The energy consumption data used in this study were collected on an hourly basis from six residential households and enriched with meteorological data, including temperature, humidity, wind, and precipitation. This granular level of detail allowed the models to learn fine-grained consumption patterns influenced by both behavioural routines and environmental conditions. However, extending the prediction horizon to five days presents a unique challenge: while the models perform well on short-term hourly forecasting, reliable multi-day forecasting requires additional inputs, particularly accurate future weather predictions and recent consumption trends.

For example, predicting energy usage 5 days ahead cannot rely solely on past energy data; it must also incorporate forecasted temperatures, since temperature fluctuations have a significant impact on heating or cooling-related energy demand. Moreover, capturing recent consumption changes from the previous 3 hours can help the model adjust dynamically to short-term behavioural or environmental shifts.

This reveals a tricky but solvable challenge: in order to predict 5 days of energy consumption, the model must be integrated into a pipeline that includes:

1. The temperature forecasting component provides hourly temperature predictions for the next 5 days.
2. Real-time consumption tracker, supplying the model with the latest usage data (e.g., last 3 hours).
3. Forecasting engine, trained with these combined features to output energy predictions per hour for the next 120 hours.

This pipeline ensures that the model does not rely solely on static historical data, but also actively adjusts based on expected external conditions and current usage trends. This paves the way for fully automated demand forecasting systems that can help with grid-level energy management and enable smarter home automation systems.

Developing an energy consumption forecasting model faces several challenges, including incomplete data from meteorological sources, noise, fluctuating usage patterns, and external factors not present in the dataset (Ruiz et al., 2020).

Although the head attention mechanisms used in Long Short-Term Memory (LSTM) improve model interpretability, they are computationally demanding and may not scale efficiently. Furthermore, the models performed well under typical consumption patterns but exhibited reduced accuracy during unexpected or unexplained spikes or dips in energy consumption. This is likely due to the limited availability of such extreme events in the training dataset (Vaswani et al., 2017).

7.3 Challenges and limitations

Predicting residential energy consumption goes far beyond estimating how much the next energy bill will cost. By accurately forecasting energy consumption, energy suppliers and distributors can more effectively balance supply and demand, reducing operating costs and optimising energy procurement strategies (Barthelmie, Murray, and Pryor, 2008).

More reliable energy consumption forecasting models enable energy companies to better respond to changes in demand, thereby minimising energy waste and in turn improving system reliability (Suganthi and Samuel, 2012).

The results of this study highlight the potential of advanced machine learning models, particularly LightGBM, to support data-driven decision-making in energy management.

This work supports the company's efforts to improve energy management, improve forecasting processes, and promote smarter resource usage.

7.4 Lessons learned

During de development of the models, it was discovered that, although household energy consumption is influenced by ambient temperature, the most significant drivers are human activities, such as appliances being run, showers being taken, or devices being used at home. These behaviour-driven spikes are often found to outweigh gradual weather-related trends. It was proven that capturing occupancy patterns and specific appliance usage is essential for improving forecast accuracy.

Developing the models, including the transformer models for regression, were also studied, and it was found that their attention mechanisms are capable of learning long-range dependencies in time series data is a valuable asset. However, careful hyperparameter tuning and sufficient training data are required by these models to avoid overfitting. In future work, the focus will be placed on combining transformer-based architectures with rich behavioural and appliance-level features to enhance multi-day energy forecasts.

In addition to its technical contributions, this study offers potential for renewing working life practices in the energy sector. Forecasting models like the ones tested here could be integrated into demand response systems used by energy companies or municipalities to anticipate and manage consumption more proactively. For example, by combining the model with user interfaces that provide real-time consumption feedback, households could make more informed decisions, leading to behavioural changes and reduced energy waste. This aligns with Finland’s energy transition goals and could support policy development, smarter billing systems, or sustainability-focused incentives based on predictive data. As such, the project offers not only technical accuracy but also a foundation for innovation in consumer engagement and regional energy planning.

8 Conclusion and future work

8.1 Answers to research questions

For the first question, the model can predict energy use for each of the next 24 hours when it is built on a robust feature pipeline implemented in Python pandas. Using three moving average windows of past consumption and their three hours standard deviations gives the model clear insight into recent usage patterns and supports accurate hour by hour forecasts (Peng et al., 2019).

Predicting energy use for each hour over the next five days is far more challenging because household routines and weather conditions can shift significantly over multi day spans. These changes introduce noise that accumulates across hours and degrades accuracy when the model tries to extend beyond one day (Sevlian & Rajagopal, 2014).

In this thesis, all preprocessing was done in Python, pandas and the only source of data leakage arose from applying scaling before splitting into training and test sets. To avoid this, the data were first split in time order, then scaling parameters were learned on the training subset and applied separately to validation and test subsets. According to Kuhn & Johnson this procedure ensures that information from future observations never influences model fitting or evaluation.

By combining a carefully designed pandas-based pipeline with correct temporal splitting and feature scaling the model delivers reliable 24 hours forecasts. Extending accuracy to five-day hourly predictions would require additional context features and a much larger and more diverse sample of households.

8.2 Recommendations for future development

Looking ahead, expanding the number of participating households would be a valuable step forward. In particular, including families that switch to real time pricing contracts, where electricity costs are directly linked to wholesale market rates, can help the model better understand how people adjust their behaviour based on price changes. By recording the exact start date of each new contract and the price per kilowatt hour, the forecasting system can more clearly detect how consumption patterns shift when prices rise or fall. (Turney, 2023).

Another meaningful improvement would be to simplify how we analyse energy use. Instead of focusing on hourly readings, we could group consumption into daily totals. This approach smooths out short term spikes caused by things like running the washing machine late at night or sudden changes in weather. Daily summaries reveal more stable and consistent patterns in energy use, leading to forecasts that are both clearer and more reliable (Investopedia, 2009).

As the dataset continues to grow, it also makes sense to transition the heavy lifting of data processing into a structured database environment using SQL. Tasks such as joining tables, filtering records, and calculating moving averages can be handled efficiently within the database. That way, the data that arrives in Python is already clean and ready for modelling. With this foundation in place, Python can focus on what it does best, finding patterns, refining algorithms, and generating accurate energy predictions (Turney, 2023).

8.3 Final thoughts

Forecasting hourly energy usage can feel like trying to hit a moving target. People don’t follow perfect schedules. One day, a person might arrive home at three in the afternoon and turn on the oven immediately. The following Monday, the same person might get stuck in traffic and come home hours later. Arriving home an hour late or turning on appliances at unexpected times can create sudden spikes or drops in consumption that can throw a model out of whack. A surprise storm can also increase demand for lighting or heating in the evening, even when the forecast had promised sunshine. This kind of sudden highlights the importance of integrating real-time the forecast predictions

Finally, there’s the sheer volume of information that’s relevant to forecasting electricity consumption. Recording and processing thousands of hourly readings from dozens or hundreds of homes requires considerable computing power. The more data you feed into the model, the harder it becomes to separate real patterns from random noise. And if households start changing their habits, like kids going on vacation or someone taking a night shift, the model has to relearn everything. In practice, this means frequent updates and constant vigilance, or these hourly forecasts can quickly divert of its course.

One way to reduce this volatility is to predict energy use at the daily level instead of hour by hour. By focusing on 24-hours aggregates the model sees broader trends and smoother patterns, often leading to more accurate forecasts (Sevlian & Rajagopal, 2014).

Furthermore, adding households to the dataset at the moment they sign new supply contracts can help capture how tariff structures affect consumption. Tracking contract initiation dates and including the corresponding energy price per kilowatt hour allows the model to learn how variations in cost influence user behaviour and overall demand. Incorporating these pricing features alongside contract events can improve the model’s ability to forecast under different market conditions. (Faruqui & Sergici, 2010).

In markets where residential tariffs are tied to wholesale spot prices, the price per kilowatt-hour becomes a key demand driver, leading households to migrate to real-time price contracts (Borenstein, 2005).

In this thesis, all participating households are under fixed-price contracts, the constant price of energy was included as a feature in the model to capture its impact on consumption patterns. If variable-price contracts were introduced in the future, it would be necessary to record the start date of each contract and the evolution of hourly prices, making the spot price an essential resource to capture the influence of volatility on demand.

Predicting energy use for every single hour can feel like chasing shadows. Aggregating consumption into 24-hour totals smooths out the random spikes and drops that come from small routine changes thereby often improving forecast accuracy (Sevlian & Rajagopal, 2014).

Another approach is to take when dealing with very large volumes of data. With huge data sets, it makes sense to let the database do the heavy lifting of combining tables, crunching numbers, and filtering out unwanted rows using SQL, rather than trying to do it all in Python. Databases excel at set-based operations because they use indexes and parallel processing to reduce preprocessing time (Hellerstein, Stonebraker, & Hamilton, 2007).

In summary, this thesis meets its research objectives by comparing three advanced forecasting methods. At the same time, it provides a foundation for improving energy awareness, operational planning, and long-term sustainability in the energy sector.

References

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Wah, T. Y. (2015). Time-series clustering – A decade review. *Information Systems, 53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007

Arshad, M. (2024). *Machine learning for smart grid systems*. Springer.

Barthelmie, R. J., Murray, F., & Pryor, S. C. (2008). The economic benefit of short-term forecasting for wind energy in the UK electricity market. *Energy Policy, 36*(5), 1687–1696. https://doi.org/10.1016/j.enpol.2008.01.027

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks, 5*(2), 157–166.

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences, 191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bishop, C. M. (2006). *Pattern recognition and machine learning*. O’Reilly Media.

Brownlee, J. (2016). Why one-hot encode data in machine learning? *Machine Learning Mastery*. Retrieved March 3, 2025, from https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/

Brownlee, J. (2018). *Deep learning for time series forecasting: Predict the future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.

Brownlee, J. (2020). *Multistep time series forecasting with LSTMs in Python*. Machine Learning Mastery. Retrieved from https://machinelearningmastery.com/multistep-time-series-forecasting-with-lstms-in-python/

Chollet, F. (2021). *Deep learning with Python* (2nd ed.). Manning Publications.

Doumèche, N., Allioux, Y., Goude, Y., & Rubrichi, S. (2023). Human spatial dynamics for electricity demand forecasting: The case of France during the 2022 energy crisis. *arXiv*. https://arxiv.org/abs/2301.09137

Eichner, A. (2023). *Energy market dynamics and the role of consumption forecasting*. Springer.

Enerdata. (2024). *Global energy trends 2024*. Retrieved from https://www.enerdata.net/publications/reports-presentations/global-energy-trends.html

European Commission. (2022). *EU energy resilience strategy*. Retrieved from https://energy.ec.europa.eu/system/files/documents/eu-energy-resilience-strategy-2022_en.pdf

Feature-Engine. (2023). *Feature-engine: Feature engineering for machine learning*. https://feature-engine.readthedocs.io/en/latest/

Finnish Meteorological Institute. (2025). *Weather warnings archive*. https://en.ilmatieteenlaitos.fi/weather-warnings

Florek, P., & Zagdański, A. (2023a). A histogram-based gradient boosting approach for energy prediction. *arXiv*. https://arxiv.org/abs/2302.10788

Florek, T., & Zagdański, K. (2023b). Improving LightGBM performance with histogram-based training. *Energy Informatics, 6*(1), Article 87. https://doi.org/10.1186/s42162-023-00376-9

Fumo, N., & Biswas, M. R. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews, 47*, 332–343. https://doi.org/10.1016/j.rser.2015.03.027

Geff, D., Vaswani, A., & Kim, Y. (2017). Long-term memory challenges in neural network forecasting. *IEEE Transactions on Neural Networks, 28*(9), 2212–2225.

Geron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly.

Glassner, B. (2021). *Applied predictive modeling*. Springer.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Hawthorne, G., & Elliott, P. (2005). Imputing missing data: strategies for longitudinal studies. *BMC Medical Research Methodology, 5*, Article 5. https://doi.org/10.1186/1471-2288-5-5

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. https://otexts.com/fpp2/

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

International Energy Agency. (2023). *Finland 2023: Energy policy review*. https://www.iea.org/reports/finland-2023

International Energy Agency. (2024). *Electricity 2024: Analysis and forecast to 2026*. https://www.iea.org/reports/electricity-2024

International Monetary Fund. (2022). *Energy crisis in Europe: How did we get here?* https://www.imf.org/en/Blogs/Articles/2022/09/16/energy-crisis-in-europe-how-did-we-get-here

Investopedia. (2009). *R-squared definition*. https://www.investopedia.com/terms/r/r-squared.asp

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Jedox. (n.d.). *Forecast accuracy: Metrics that matter*. https://www.jedox.com/en/resources/

Johnkutty, M., & Davis, R. (2024). *Mastering data profiling for machine learning*. Packt Publishing.

Joseph, A. (2022). *Time series analysis and forecasting for machine learning*. Wiley.

Junninen, H., et al. (2004). Methods for imputation of missing values in air quality data sets. *Atmospheric Environment, 38*(18), 2895–2907.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Li, Y. (2024). *Time series forecasting with advanced machine learning models*. Springer.

Loshin, D. (2006). *Data quality: The accuracy dimension*. Morgan Kaufmann.

Lu, S., Yuan, W., Schwartz, P., & Benjamin, S. (2007). Lag-based data enhancement for energy load prediction. *IEEE Power Engineering Society General Meeting*, 1–6.

Mahajan, D., Singh, A., & Bruns, C. (n.d.). Encoding cyclical features for time series forecasting. *Medium*. https://medium.com/encoding-cyclic-features

Mount, C. (2021). *Exploratory data analysis in practice*. O'Reilly.

Munir, S., Pradhan, M. R., Abbas, S., & Khan, M. A. (2023). Energy consumption prediction based on LightGBM empowered with explainable artificial intelligence. *IEEE Access, 11*, 112001–112011. https://doi.org/10.1109/ACCESS.2023.3327760

Oliveira, G. M., & Oliveira, R. C. L. (2023). Transformers for energy forecast. *Sensors, 23*(15), 6840. https://doi.org/10.3390/s23156840

Olson, J. E. (2003). *Data quality: The accuracy dimension*. Morgan Kaufmann.

OpenAI. (2025). *ChatGPT (Mar 14 version) [Large language model]*. https://chat.openai.com/chat

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

ResearchGate. (n.d.). Electricity consumption prediction based on LSTM with attention mechanism. https://www.researchgate.net/publication/338424977

Ruiz, P. A., Makarov, Y. V., Huang, Z., & Etingov, P. V. (2020). Addressing uncertainty and variability of renewable generation with robust and stochastic optimization. *IEEE Transactions on Power Systems, 35*(1), 455–465. https://doi.org/10.1109/TPWRS.2019.2927653

Sarka, D. (2021). *Data preparation using SQL and Excel*. Apress.

Sciences Force. (2024). Hybrid attention-enhanced deep learning for hourly energy forecasting. *International Journal of Smart and Wireless Applications, 6*(1), Article 314. https://sciencesforce.com/index.php/iswa/article/download/314/483/847

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications*. Springer.

Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2019). A comparison of ARIMA and LSTM in forecasting time series. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1394–1401.

Skforecast. (2023). *Cyclical encoding of time features*. https://skforecast.org/latest

Soltani, A., & Meinke, H. (2004). Linear interpolation for missing climate data. *Environmental Modelling & Software, 19*(5), 503–509.

Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting – A review. *Renewable and Sustainable Energy Reviews, 16*(2), 1223–1240.

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews, 13*(8), 1819–1835.

Tabasi, M., & Forotan, F. (2016). Regression analysis on electricity consumption based on weather conditions. *Energy Economics, 58*, 174–181.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting, 16*(4), 437–450.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.

Wang, Y., Nie, W., Cao, D., & Liu, Y. (2023). Multi-head attention models for time-series energy forecasting. *Energy AI, 12*, 100153.

Wu, H., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems, 34*, 22419–22430.

Zatzman, G. M. (2012). *Sustainable energy pricing*. John Wiley & Sons.

Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., & Eickhoff, C. (2021). A transformer-based framework for multivariate time series representation learning. *Proceedings of the 27th ACM SIGKDD Conference*, 2114–2124. https://doi.org/10.1145/3447548.3467401

Zhang, G., Eddy Patuwo, B., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35–62.

Appendices

Appendix 1. Full Project Code

Find my codes at GitHub accessing at: https://github.com/sarakfouri/master_thesis