PREDICTING ELECTRICITY CONSUMPTION IN NORWAY:

A COMPARISON OF MACHINE LEARNING MODELS

Abstract

This paper proposes a model for the estimation of the consumption of electricity in Norway, which can accurately predict the next 24 h of load with and estimation of load for 1 week to 1 month using in one-hour intervals. It shows the utilization of one timeseries model; ARIMA, two regression models; Linear regression and Ridge Regression Models, two Ensemble; XGBoost and Random Forest, and one Neural network; MLP. In addition, the present paper shows the way to significantly improve the accuracy of the prediction through feature engineering, ensemble machine learning and neural network process.

Upon rigorous evaluation, the ARIMA model demonstrated exceptional predictive accuracy across all zones, as evidenced by the close alignment between the actual and predicted values of electricity consumption. Furthermore, meticulous parameter tuning, and model optimization techniques were employed to enhance the model's predictive performance and generalization capabilities.

The rest of the models performed equally well with XGBoost leading with 99% accurate predictions.

Ensemble methods like XGBoost and Random Forest excel in capturing complex relationships and achieving high accuracy, linear regression-based approaches offer simplicity and interpretability. MLPs provide a flexible and powerful modelling framework that can capture intricate patterns in the data. Understanding the trade-offs and strengths of each algorithm is crucial for making informed decisions in predictive modelling tasks, especially in domains like energy consumption forecasting where accurate predictions are essential for effective resource management and decision-making.

In conclusion, the findings of this study highlight the efficacy of ARIMA as a valuable tool for forecasting electricity consumption across diverse geographical zones.

Table of Contents

Chapter 1: Introduction

Motivation

Electricity consumption is an essential indicator of a country’s development and economic growth (Vosooghzadeh, 2021). The reliable and efficient supply of electricity is a critical component of modern societies, supporting economic growth, technological advancement, and overall quality of life.

In an era marked by rapid urbanization, industrialization, and a growing global concern for sustainable energy management, accurately forecasting electricity consumption has emerged as a pivotal challenge (Wang et al., 2021). Data-driven solutions are transforming the way we approach complex problems, predicting electricity consumption stands at the crossroads of technological innovation and sustainable energy management (Banik et al., 2021).

The Kingdom of Norway, renowned for its commitment to green energy sources and an enviable wealth of hydropower resources, stands as a beacon of clean and reliable energy generation (Norby et al., 2019). The need for accurate electricity consumption forecasts in Norway is underscored by the unique characteristics of its power generation and consumption patterns, influenced by factors such as weather, seasonality, and energy policy.

Goal

The intention of this thesis is to employ machine learning and neural network methods to make prediction of power consumption one day ahead in 15 minutes interval at country scale in Norway.

Norway's distinctive energy profile, shaped by its geographical diversity, climate variations, and renewable energy policies, presents a complex and dynamic backdrop for forecasting. Traditional approaches to electricity consumption prediction may struggle to capture the nuances and ever-evolving factors affecting Norway's energy landscape.

By leveraging historical consumption data, weather variables, economic indicators, and other relevant features, machine learning algorithms and neural networks can be trained to generate forecasts that account for the intricacies of Norway's energy system.

This paper proposes a model for the estimation of the consumption of electricity in Norway, which can accurately predict the next 24 h of load with and estimation of load for 1 week to 1 month using in 15-minute intervals. In addition, the present paper shows the way to significantly improve the accuracy of the prediction through feature engineering, ensemble, hybrid machine learning and neural network process. The analyses or findings also provide interesting results in connection with energy consumption.

Thesis structure

In chapter 1, an introduction is given about the motivation and the goal in this research. In chapter 2, the main background concepts in this research are presented, including Norwegian energy landscape, factors affecting electricity consumption, and machine learning.

In chapter 3, there is a review on the relevant studies about machine learning models and neural works, including using machine learning/neural networks on time series data, using machine learning neural network on power consumption, and using an ensemble of machine learning models on power consumption.

In chapter 4, there will be a discussion of the research framework for gathering the data, how the data is analyzed, and a further explanation of why specific methods and models used in the analysis were chosen. Limitations will also be mentioned. Machine learning and artificial neural networks are introduced briefly. This includes their basic components, learning processes and the several variants that are picked out for this research.

In chapter 5, it will outline the findings of our analysis and models. We will briefly explain the facts and results from our models and present illustrations to visualize these results using graphs and tables. After the results are presented, interpretation of the results will be discussed concerning the objective of the thesis.

In chapter 6, we will highlight what we have achieved. The limitations mentioned in the methods chapter will also be considered to better ground our thesis. Our experiences and suggestions for further research will end this bachelor thesis.

Chapter 2: Background

Electricity is a commodity that can be bought, sold, or traded and is usually measured in Kilowatt hour (Chicago Mercantile Exchange (CME) Group). However, it is different from the other commodities such as oil and gas. Electricity has a different value over space because its flows cannot be controlled easily and efficiently, and transmission components must be operated under safe flow limits. If not, there is a risk of cascading failures. Secondly, electricity generation must always meet demand and to achieve this, it should be flexible. Finally, electricity has a different value over time, which means large volumes cannot be stored economically yet (Chicago Mercantile Exchange (CME) Group). These three unique characteristics are the reason Norway is connected to several electricity markets.

In this chapter, we will discuss Norway's Energy landscape and the factors affecting electricity consumption in Norway.

Electricity in Norway

Norway, renowned for its abundant fjords and lakes, relies significantly on its natural resources in its electricity market. The country's primary energy sources are renewable, particularly hydropower, with a minor contribution from wind energy (Egging & Tomasgard, 2018) and solar energy. Even though Norway is a major oil producer, fossil fuels account for less than two percent of Norway's electricity generation. As a result, Norway has become one of the world's top per capita electricity consumers, thanks to its reliable power supply (Statista, 2020).

The country has a long history of using hydropower, dating back to the late 19th century. Hydropower is the cornerstone of Norway's electricity production, generating an average of 140 terawatt-hours yearly. Presently, over 90 percent of Norway's electricity is generated by more than 1,700 hydropower plants (IEA, 2022) (Egging & Tomasgard, 2018). Norway leads in hydroelectricity production in Europe and globally. Its hydropower capacity has steadily increased for over a decade, and much of it is flexible, with storage reservoirs controlled by dams. Over a thousand such reservoirs are scattered throughout the country, enabling better supply-demand balance year-round (Statista, 2020).

In 2020, Norway's electricity consumption reached approximately 126 terawatt-hours, and peaked in 2021 at 139.5 terawatt-hours and decreased to 133.4 terawatt-hours in 2022 with private households being the largest consumers, using electricity for lighting, heating, and appliances. Power-intensive manufacturing, like aluminum production, ranked second in consumption. Consequently, Norway had a per capita electricity consumption of 26 megawatthours in 2020, more than double that of the United States that year. Furthermore, Norway boasts some of the lowest average household electricity prices in Europe, partly due to its renewable energy sources (Statista, 2020).

As a leader in renewable energy production, Norway experiences fluctuations in electricity generation due to weather conditions and seasonal variations. Hydroelectric power production is heavily dependent on the water supply, which can vary significantly between seasons. Norway's unique climate, with cold winters and milder summers, results in distinct electricity consumption patterns. Winter months typically see increased demand due to electric heating requirements, while summer months may have lower consumption (Madeline et al.). Additionally, the country experiences significant seasonal daylight variations, impacting the need for artificial lighting.

Norway's electricity market is divided into five bidding zones (Statnett), each with its specific characteristics and electricity consumption patterns (NordPool).

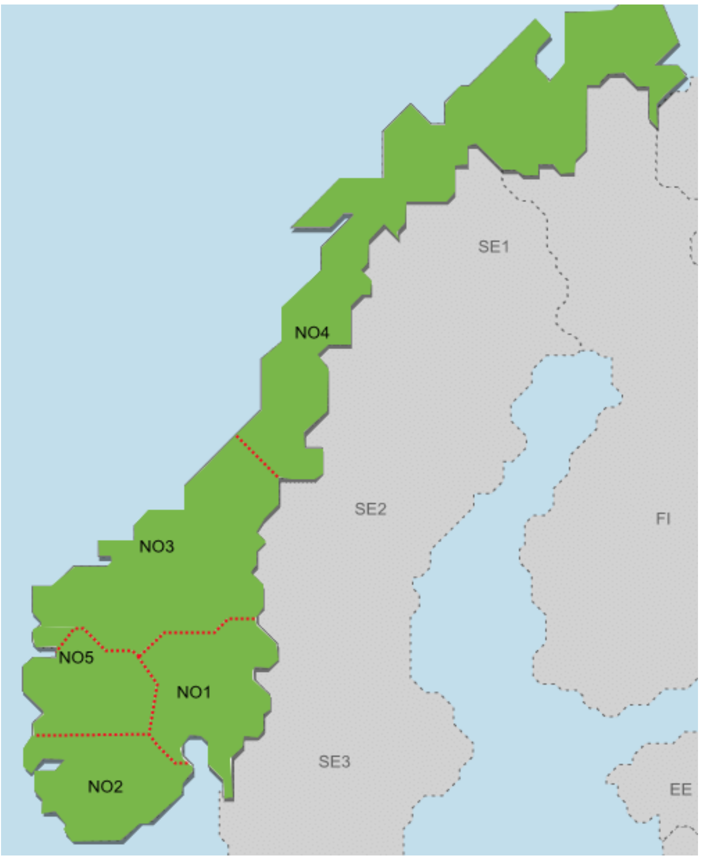


Fig 1:A figure representing the power situation in Norway.

These bidding zones—Southern Norway; Kristiansand(NO2), Western Norway; Bergen(NO5), Central Norway; Oslo(NO1), Northern Norway; Mølde & Trondheim(NO3), and Tromsø(NO4)—differ in terms of their geographical locations, climate, and industrial activities (NVE, 2016). Accurate predictions are vital for managing electricity generation and distribution efficiently across these zones.

Accurate forecasting of electricity consumption plays a vital role in the reliable operation of the energy grid and facilitates effective energy resource management, pricing, and grid stability. Led by the state-owned transmission systems operator (Tso & Yau), Statnett, the electricity market in Norway has ample interconnection capacity and an organised emergency preparedness system. Norway has been actively advancing its energy infrastructure with a focus on smart grids and smart meters to optimize electricity consumption. The implementation of these technologies aligns with Norway's commitment to sustainability, efficient energy use, and the integration of renewable energy sources.

Smart grids in Norway leverage advanced communication and control systems to enhance the flexibility, reliability, and efficiency of the electricity grid. These grids enable better management of electricity consumption by facilitating real-time communication between consumers and utility providers. They also support the integration of renewable energy sources by efficiently balancing the intermittent nature of sources like wind and solar.(Heba et al., 2023)

Smart meters play a pivotal role in this transformation by providing consumers with detailed and real-time information about their electricity usage. With the ability to transmit data remotely, smart meters eliminate the need for manual readings and enable more accurate billing based on actual consumption. Consumers can access this information through online portals or mobile apps, gaining insights into their energy usage patterns, peak consumption times, and overall efficiency.

The implementation of smart grids and meters in Norway contributes to demand-side management strategies (NVE). Consumers can make informed decisions about when to use electricity based on pricing variations throughout the day, promoting load balancing and reducing peak demand. This capability is particularly important as it enhances the grid's overall stability and efficiency (Heba et al., 2023).

Factors affecting electricity consumption in Norway.

Research is needed to improve the accuracy of short-term electricity consumption forecasts, considering factors such as population growth, technological advancements, and changes in energy policies over extended periods. Given Norway's unique energy landscape, this master thesis aims to compare machine learning models and neural networks in predicting day-ahead electricity consumption, considering several factors specific to the country's energy ecosystem.

According to (Demanuele C. et al., 2010) one of the main challenges of energy forecasting is the highly variable and unpredictable nature of occupant behavior. However, the variables controlled by the occupant are one of the most influential factors in determining energy consumption. Other factors that can influence the accuracy of energy consumption prediction models include environmental factors such as temperature, humidity, wind strength, and rain, as well as holidays and weekends. (Khan et al., 2021) also mentions that researchers have analyzed factors such as temperature changes and energy use during hurricanes to explore critical errors.

I. Weather Conditions

Weather is a crucial factor influencing electricity consumption. Temperature, humidity, wind speed, and sunlight hours affect heating, cooling, and lighting needs, thereby impacting electricity usage. For example, hot weather increases the demand for air conditioning, while cold weather increases the demand for heating. (Moral-Carcedo & Vicens-Otero, 2005)

Norway's electricity consumption is significantly influenced by its weather conditions, particularly due to its reliance on hydropower. Hydroelectric plants generate most of Norway's electricity, making them highly susceptible to changes in precipitation and temperature. In periods of heavy rainfall or melting snow, hydroelectric reservoirs experience increased water inflow, boosting electricity production. Conversely, during droughts or prolonged cold spells, water levels may decrease, leading to decreased electricity generation and potentially necessitating alternative energy sources. This dependence on weather patterns underscores the importance of meteorological forecasting for energy planning in Norway. (Madeline et al., 2022)

Weather conditions and seasonal changes exert a significant influence on Norway's electricity consumption due to its heavy reliance on hydropower. Precipitation patterns, particularly rainfall and snowmelt, directly impact the water levels in reservoirs, thereby affecting hydropower generation. During periods of abundant rainfall or snowmelt, such as in the spring and fall, reservoirs experience higher water inflow, leading to increased electricity production. Conversely, dry spells or exceptionally cold temperatures during winter can result in reduced water levels and decreased hydropower output, necessitating alternative energy sources to meet demand.

Seasonal changes also play a crucial role in Norway's electricity consumption patterns. Winter months typically see higher energy usage due to increased heating needs, particularly for electric heating systems prevalent in many Norwegian households. Additionally, shorter daylight hours during winter necessitate more artificial lighting, further contributing to electricity demand. Conversely, during summer months, electricity consumption may decrease as heating requirements diminish and natural lighting reduces reliance on artificial sources. (Madeline et al., 2022)

II. Time of Day and Day of Week

Electricity consumption patterns vary throughout the day and week. Peaks typically occur during weekday afternoons and evenings when commercial and residential activities are at their highest. Weekends and holidays often have different consumption patterns.

The time of day and day of the week significantly impact Norway's electricity consumption patterns, reflecting various factors such as industrial activity, commercial operations, and residential usage. During weekdays, electricity demand typically peaks in the morning and evening hours as people engage in activities like cooking, heating, and using electronic devices. Industrial and commercial sectors also contribute to this peak demand during regular working hours. Conversely, electricity consumption tends to decrease during nighttime and weekends when industrial activity is lower, and many businesses are closed.

The variability in electricity consumption throughout the day and week underscores the importance of load forecasting and grid management to ensure a reliable supply of electricity. Norway's grid operators rely on sophisticated modelling and forecasting techniques to anticipate fluctuations in demand and adjust energy production, accordingly, minimizing the risk of power shortages or overloads.

III. Technological Advances

The adoption of new technologies, such as energy-efficient appliances, smart meters, and renewable energy sources, can alter electricity consumption patterns. Incorporating information about technological advancements helps improve prediction accuracy.

The adoption of energy-efficient technologies, such as LED lighting, smart appliances, and electric vehicles (Gievska & Madjarov), has led to reduced electricity consumption in various sectors. Energy-efficient appliances and lighting systems consume less power while providing the same level of service, contributing to overall energy savings. Similarly, the increasing popularity of EVs has led to a higher demand for electricity but also offers opportunities for demand-side management through smart charging solutions, which can help balance grid loads and optimize energy usage.

Moreover, advancements in renewable energy technologies, particularly in solar and wind power, have expanded Norway's electricity generation capacity beyond traditional hydropower. The integration of these intermittent renewable sources into the grid requires innovative solutions such as energy storage systems and grid management technologies to ensure grid stability and reliability. Additionally, advancements in grid infrastructure, such as smart meters and digital control systems, enable more efficient monitoring and management of electricity distribution, contributing to overall system efficiency and reliability.

IV. Policy and Regulation

Government policies, regulations, and incentives related to energy conservation, renewable energy integration, and carbon emissions can influence electricity consumption patterns. Changes in policy can lead to shifts in consumer behaviour and energy usage.

Policy and regulation wield considerable influence over Norway's electricity consumption dynamics, intricately shaping both demand and supply aspects of the energy sector. One significant facet of policy intervention lies in the realm of energy efficiency standards mandated by government regulations. For instance, stringent building codes necessitating the use of energy-efficient appliances and robust insulation techniques translate into tangible reductions in electricity consumption across residential, commercial, and industrial sectors. By setting minimum efficiency requirements, such policies not only mitigate energy waste but also foster a culture of sustainability among consumers and businesses alike.

Moreover, regulatory frameworks aimed at curbing carbon emissions and fostering renewable energy deployment constitute pivotal drivers of Norway's electricity consumption landscape. Measures such as feed-in tariffs and renewable energy targets incentivize investments in clean energy technologies, thereby progressively steering consumption patterns towards more sustainable sources. This not only aligns with broader environmental objectives but also contributes to the diversification and resilience of the energy mix, diminishing dependence on finite fossil fuel resources.

In addition to promoting energy efficiency and renewable energy adoption, policy instruments often encompass demand-side management strategies tailored to optimize consumption patterns. Time-of-use pricing schemes and peak demand reduction incentives encourage consumers to adjust their electricity usage in response to fluctuating demand profiles, thereby enhancing overall grid stability and efficiency. Such initiatives empower consumers to make informed decisions regarding their energy consumption habits, fostering a more responsive and adaptive energy ecosystem.

Furthermore, the electrification of transportation emerges as a key focal point within Norway's energy policy landscape. Government support through subsidies for electric vehicles (Gievska & Madjarov) and the establishment of charging infrastructure networks not only spurs demand for electricity but also facilitates the transition towards cleaner and more sustainable mobility solutions. By encouraging the adoption of EVs, policymakers aim to concurrently bolster electricity consumption while mitigating the environmental impact associated with conventional fossil fuel-powered vehicles.

Overall, the intricate interplay between policy and regulation serves as a cornerstone in steering Norway's electricity consumption trajectory towards greater sustainability, resilience, and efficiency. Through a multifaceted approach encompassing energy efficiency mandates, renewable energy incentives, demand-side management initiatives, and support for electric vehicle adoption, policymakers endeavour to foster a robust and sustainable energy ecosystem that aligns with national and global sustainability objectives.

V. Bidding Zones

Norway's five bidding zones, each with its unique energy policies, exert a profound influence on electricity consumption patterns across the nation. In the NO1 zone, which encompasses urban centers like Oslo, policies promoting renewable energy sources and energy efficiency measures directly impact consumption trends. Increased adoption of energy-efficient technologies in residential, commercial, and industrial sectors reduces overall electricity demand, mitigating consumption growth despite population and economic expansion. Similarly, the emphasis on renewable energy sources such as hydropower ensures a reliable and sustainable electricity supply, encouraging consumers to utilize cleaner and more environmentally friendly energy options.

Transitioning to the NO2 bidding zone, characterized by coastal regions like Bergen and Stavanger, policies favoring the development of wind and wave energy resources influence electricity consumption dynamics. The integration of offshore wind farms and marine energy projects diversifies the energy mix, reducing dependence on conventional fossil fuels and lowering greenhouse gas emissions. As a result, shifts in energy production towards renewables translate into shifts in consumption patterns, with consumers incentivized to embrace greener energy alternatives and reduce their carbon footprint.

In the NO3 zone, comprising areas like Kongsvinger and Østfold in the eastern part of Norway, grid modernization and integration efforts play a pivotal role in shaping electricity consumption trends. Upgrades to the electricity grid infrastructure enhance reliability and resilience, reducing the frequency and duration of power outages and minimizing disruptions to consumers. Additionally, the seamless integration of renewable energy sources into the grid facilitates the uptake of distributed generation systems, enabling consumers to generate their electricity locally and participate in demand-side management initiatives. This, in turn, leads to more efficient utilization of electricity resources and greater flexibility in consumption patterns.

Heading north to the NO4 bidding zone, encompassing remote regions like Tromsø and Bodø, energy policies focused on decentralization and energy self-sufficiency influence electricity consumption behaviors. Initiatives promoting microgrids, community-based renewable energy projects, and off-grid solutions empower local communities to take control of their energy supply, reducing reliance on centralized electricity generation and transmission infrastructure. By fostering a culture of energy independence, consumers in the NO4 zone are incentivized to adopt energy-efficient practices and embrace renewable energy technologies, leading to more sustainable consumption patterns and reduced reliance on imported electricity.

Finally, in the NO5 bidding zone, which includes regions like Trondheim and Lillehammer in central Norway, market liberalization and integration efforts impact electricity consumption through increased competition and market efficiency. Policies promoting market transparency, grid connectivity with neighboring countries, and cross-border energy trading create opportunities for consumers to access a diverse range of energy sources and services. This fosters innovation and drives efficiency improvements throughout the energy value chain, ultimately leading to more optimized consumption patterns and enhanced consumer choice.

The energy policies of Norway's bidding zones have a significant influence on electricity consumption patterns, shaping consumer behavior, driving technological innovation, and fostering a transition towards a more sustainable and resilient energy system. By aligning policy objectives with environmental sustainability, energy security, and consumer welfare, Norway continues to chart a course towards a low-carbon future, where electricity consumption is not only efficient and reliable but also environmentally sustainable and socially equitable.

Machine Learning and Neural networks for electricity consumption prediction

Machine learning and neural networks have emerged as powerful tools for electricity consumption prediction, offering unparalleled accuracy and flexibility in modeling complex consumption patterns. By leveraging historical consumption data, weather variables, demographic information, and other relevant factors, machine learning algorithms can identify hidden patterns and relationships to make accurate predictions of future electricity demand. Neural networks, in particular, excel in capturing intricate nonlinear relationships within the data, making them well-suited for forecasting tasks where traditional linear models may fall short.

One of the key advantages of machine learning and neural networks in electricity consumption prediction is their ability to handle large volumes of data efficiently. With the proliferation of smart meters and advanced sensor technologies, utilities have access to vast amounts of data on electricity consumption, spanning different time intervals and geographic regions. Machine learning algorithms can effectively process and analyze this data to extract valuable insights and generate accurate predictions of future consumption trends.

Moreover, machine learning models offer a high degree of flexibility and adaptability, allowing them to capture complex dependencies and nonlinear relationships that may exist between various factors influencing electricity consumption. For example, neural networks can automatically learn and adjust their internal parameters to optimize performance based on the characteristics of the input data, making them well-suited for modeling dynamic and evolving consumption patterns.

In addition to their predictive capabilities, machine learning algorithms also offer valuable insights into the drivers of electricity consumption, helping utilities and policymakers better understand the factors influencing demand. By analyzing the feature importance and contribution of different variables to the prediction model, stakeholders can identify opportunities for demand-side management, energy efficiency initiatives, and targeted interventions to optimize consumption patterns and reduce overall energy costs.

Furthermore, machine learning and neural networks enable real-time forecasting of electricity consumption, allowing utilities to make timely decisions and adjust supply accordingly to meet demand. This is particularly valuable in the context of renewable energy integration, where accurate forecasting of electricity demand is essential for optimizing the operation of renewable energy sources such as solar and wind power.

However, despite their numerous advantages, machine learning and neural networks also present certain challenges and considerations in the context of electricity consumption prediction. One challenge is the need for high-quality data, as the accuracy of machine learning models heavily depends on the quality and relevance of the input data. Ensuring data accuracy, consistency, and completeness is therefore crucial for obtaining reliable predictions.

Another consideration is the complexity of model interpretation and explainability, especially in the case of deep neural networks. While neural networks may offer superior predictive performance, their black-box nature can make it difficult to understand the underlying mechanisms driving predictions. This poses challenges for stakeholders seeking to interpret and trust the results produced by these models.

Chapter 3: Literature review

Accurately predicting short-term electricity consumption is a multifaceted endeavour that is critical for maintaining the stability of power grids, facilitating efficient energy trading, and effectively managing demand-side resources. In recent years, the intersection of neural networks and machine-learning techniques has emerged as a promising avenue for improving the accuracy and efficiency of such predictions. This comprehensive literature review delves deeply into various studies to evaluate the performance of these methods in short-term electricity consumption prediction, exploring their strengths, limitations, and implications.

Evaluating Methodologies: A Review of Key Studies

Gang Chen et al.: Machine-Learning-Based Electric Power Forecasting

Several studies have assessed the performance of neural networks and machine learning algorithms for electricity consumption prediction. Gang Chen et al. introduced a comprehensive framework that integrates machine learning techniques to enhance regional electricity demand forecasting. Their study, focused on Guangdong Province, China, employed linear regression, the random forest algorithm, and support vector regression (SVR) models. The SVR model demonstrated high accuracy in predicting electricity demand, emphasizing the importance of socioeconomic development and weather variability in influencing the consumption patterns. (Gang et al., 2023)

Badar and Shams: Short-Term Electrical Load Demand Forecasting Based on LSTM and RNN Deep Neural Networks

Badar and Shams conducted a detailed investigation into short-term electrical-load demand forecasting using LSTM and RNN deep neural networks. Their study, which encompassed day- and week-ahead predictions, highlighted the promise of deep learning architectures in improving forecasting accuracy, especially in capturing seasonal variations and nonlinear relationships. The research emphasized the significance of appropriate model selection and feature engineering for enhancing the prediction performance. The model introduced in this study specifically aims at achieving precise load demand prediction for smart grids, which is essential for efficient energy management. Through experimentation, the researchers examined various activation functions, ultimately finding that the DNN with leaky ReLu yielded the most accurate forecasts. Moreover, the study underscores the critical role of feature selection and output metrics, such as MAPE, MSE, and RMSE, in the assessment of forecast models. It also highlights the substantial influence of seasonal factors, such as temperature variations and working/nonworking days, on the load demand prediction accuracy in power systems. (Badar & Shams, 2022)

Madeline et al.: Comparative Analysis of Artificial Intelligence Methodologies and Metaheuristic Strategies

Madeline et al. (2022) conducted a comparative analysis of artificial intelligence methodologies and metaheuristic strategies for predicting monthly electricity consumption across seven countries: Norway, Switzerland, Malaysia, Egypt, Algeria, Bulgaria, and Kenya. The study, which encompassed diverse models, including artificial neural networks, adaptive neuro-fuzzy inference systems, least squares support vector machines, and fuzzy time series, revealed varying levels of accuracy and robustness among different approaches. Notably, the fuzzy time-series model demonstrated superior performance in handling limited datasets and capturing uncertainties, highlighting its potential for short-term forecasting tasks. It showed AFEs below 6% owing to its adeptness in handling limited datasets and capturing uncertainties in the data. However, the ANN model demanded extensive time-series data to generate precise results, whereas the ANFIS model excelled particularly well when dealing with unpredictable data patterns. In contrast, the LSSVM model encountered difficulties in achieving high accuracy levels owing to its lack of sparsity and sensitivity to parameters related to kernel functions, thus posing challenges in effectively training the model. This research also focused on mapping the intricate relationship between electricity consumption patterns and the various influencing factors that come into play. (Madeline et al., 2022)

Wang et al.: City-Scale Daily Electricity Consumption Forecasting

Wang et al. explored city-scale daily electricity consumption forecasting with a focus on understanding the impacts of factors such as extreme weather events and public health crises. Their study utilized data-driven models to explore the decomposition of time-series data and accurately forecast city-level energy consumption in three US metropolitan areas: Sacramento, Los Angeles, and New York. The content discusses different modelling approaches, including linear models: ASHRAE’s five-parameter piecewise linear regression model and the Heating Cooling Degree Hour model; machine learning models for time-series data: a de-composed time series model; and machine learning models for tabular data: gradient boosting tree algorithms (GBM), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (NN). (Wang et al., 2021)

The study highlights the significant influence of weather-sensitive components on daily electricity usage, with a potential 30%-50% impact, and the observed 10% reduction in electricity demand during the COVID-19 pandemic in April. Additionally, every degree of Celsius increase in summer leads to approximately 5% more daily electricity usage compared with the base load in the three metropolitan areas. Gradient Boosting Machine (GBM) provided the most accurate results with a coefficient of variation of the root mean square error (CVRMSE) less than 10%. The other machine learning models (Random Forest, Support Vector Machine, and Neural Network) underestimated the electricity usage of New York in the summer of 2018. However, the research also raised concerns regarding potential oversimplification of city-scale consumption dynamics and the need for broader considerations such as climate change and evolving consumer behaviors.

Kang and Reiner: Effect of weather on electricity consumption

The relationship between weather conditions and electricity consumption has been a topic of interest in energy research. Various studies have explored the impact of weather variables on predicting electricity demand at different levels, ranging from grid-level estimations to individual household behaviors. Recent research by Kang and Reiner (2021) delved into the impact of weather variables on residential electricity consumption in Ireland using high-resolution smart metering data. Their study employed fixed-effects models on half-hourly electricity consumption data from Irish households to demonstrate the varying sensitivities of temperature, rain, and sunshine duration on electricity demand across different periods of the day.

The findings highlighted the potential of weather factors to influence individual behavior and daily routines, offering valuable insights for understanding household energy consumption patterns. However, it underscores the importance of considering weather variables in predicting electricity consumption, both at the macro level for grid estimations and at the micro level for individual households. Understanding the complex interplay between weather conditions and energy demand is essential for developing effective energy management strategies and promoting sustainable consumption practices. (Kang & Reiner, 2021)

Khan et al.: Influencing factors on Energy Consumption Patterns

Khan et al. investigated the effects of holidays, weekends, weather, and special days on the accuracy of energy consumption forecasting. The study uses error curve training and a hybrid model with ML algorithms such as CatBoost and XGBoost to predict energy consumption. Factors that affect electricity usage comprise operational challenges, occupant behavior, weather conditions, holidays, weekends, smart grid technologies, and building design and construction.

It was discovered that holidays and weekends have a significant impact on energy consumption patterns, as during these times there are often fluctuations in the demand for electricity and overall power usage. Some of the energy consumption patterns include the surge in demand associated with heightened travel activities, increased electronic device usage, and leisure and recreational activities during holiday periods. Similarly, changes in weather conditions play a crucial role in influencing the accuracy of energy prediction models, as variations in temperature, humidity, and precipitation can lead to shifts in energy consumption behavior.

The research emphasizes the application of machine learning models to enhance the accuracy of energy consumption prediction by stressing the significance of feature selection and error analysis. Moreover, the integration of hybrid machine learning models has shown to enhance the precision and reliability of energy consumption predictions by combining the strengths of different algorithms and methodologies to better capture the complex and dynamic nature of energy usage patterns. (Khan et al., 2021)

While both neural networks and machine learning techniques offer viable solutions for short-term electricity consumption prediction, each approach possesses unique strengths and limitations. Neural networks, with their ability to capture complex temporal patterns and nonlinear relationships, excel in dynamic and volatile environments, making them well-suited for short-term forecasting tasks. Conversely, machine learning algorithms provide interpretable models with efficient computational performance, offering competitive accuracy and scalability.

Overall, neural networks and machine learning techniques offer promising solutions for short-term electricity consumption prediction, several methodological considerations and challenges must be addressed. These include model selection, feature engineering, data preprocessing, hyperparameter tuning, and the interpretation of results. Moreover, the dynamic and uncertain nature of energy systems, coupled with the influence of external factors such as weather, socio-economic trends, and policy changes, pose significant challenges to accurate forecasting. Key studies highlighted in this review demonstrate the efficacy of these methods in improving forecast accuracy and understanding consumption patterns. However, methodological challenges and the need for broader contextual considerations underscore the importance of continued research and innovation in this field.

For this reason, this study will explore a model for the estimation of the consumption of electricity in Norway, which can accurately predict the next 24 h of load with and estimation of load for 1 week to 1 month using 15-minute intervals. In addition, the present paper shows the way to significantly improve the accuracy of the prediction through feature engineering, ensemble, hybrid machine learning and neural network process. The analyses or findings also provide interesting results in connection with energy consumption.

Chapter 4: Methodology

In this section, we present the research methods used in this paper. The section also presents a description of the countries studied as well as the utilized data sources, error metrics and forecasting procedures.

The Data

Access to data is incredibly challenging and most of the electricity providers are keeping their data close to the chest. Our ambition is to compare the best performing machine learning and Deep learning models based on the Norwegian historical aggregate data. We want to use this historical data to investigate the influence of extreme weather conditions, daily, hourly, seasonal, and national holiday patterns in Norway. We will investigate the energy portfolio of each zone in Norway and any special policies.
The zones studied in this paper were Southern Norway; Kristiansand (NO2), Western Norway; Bergen (NO5), Central Norway; Oslo (NO1), Northern Norway; Molde & Trondheim(NO3), and Tromsø(NO4). The selection of these specific countries was mainly based on bidding zones according to the local Transmission System Operator (TSO).

In this research, the electricity consumption data for the five bidding zones studied were required to train and test the forecasting models. The electricity consumption data were obtained from:

https://www.entsoe.eu/ - ENTSO-E, the European Network of Transmission System Operators for Electricity, which collects the data from 39 members of TSO – Transmission system operators in 35 countries. The electricity grid in Europe is the largest interconnected electrical grid in the world.

https://frost.met.no/index.html - The Frost API provides free access to MET Norway's archive of historical measurements of temperature, precipitation, or wind as well as and climate data, available daily monthly and annually reports.

A dataset from the years Jan 2015 to April 2024 was used as the training set, while a data set for the year 2016 was used as the testing set. There was a total of 81,769 observation samples and 16 columns with no missing values.

The software used was Python and MS Excel.

Error Metrics

Several error metrics are commonly employed to evaluate the accuracy of forecasting models, including mean absolute error (MAE), mean absolute deviation (MAD), root mean square error (RMSE), average forecasting error (AFE), mean absolute percentage error (MAPE), and performance parameter (PP). In our study, we assessed the effectiveness of predictive models using three primary error metrics: root mean square error (RMSE), and R Squared (R2). Here's a breakdown of each metric:

Mean Absolute Percentage Error (MAPE):

MAPE measures the relative forecasting error as a percentage of the actual value. It is calculated as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

Root Mean Square Error (RMSE):

RMSE provides an overall assessment of accuracy by considering the squared differences between predicted and actual values. The formula for RMSE is:

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

Mean Absolute Error (MAE):

MAE represents the average absolute difference between predicted and actual values. It is computed as:

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

These metrics help us assess the performance of our models, with RMSE being particularly popular in regression problems.

Descriptive Analysis and Data Preprocessing

Effect of Temperature

It is noticed that the electrical load demand increases with the decline in temperature during the winter season and it decreases in the warm season. Therefore, the seasonal variables should be included in the predicted model input to obtain accurate predictive results. A review of the literature shows that there is a strong correlation between seasonal variables and load demand. There is more need for electricity when the temperature falls below 10°C due to heating requirements in a family.

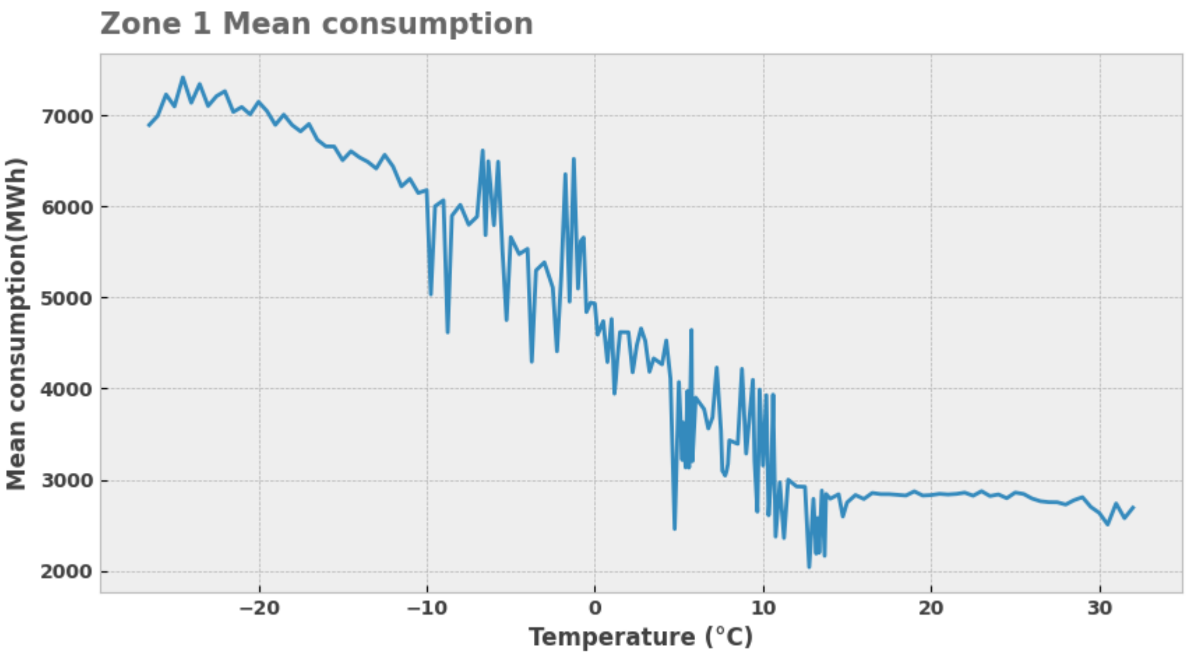


Fig 2:A figure illustrating the effect of temperature on electricity demand in Norway.

In the winter 2023-2024 was considered a harsh winter especially in zone 1 with temperature going as low as 26 °C. This means more heating was required hence high electricity demand.

Working and Nonworking Days.

Electricity usage is higher on weekdays while electricity consumption is low on Saturday and Sunday, and on other public holidays. The “Workday”, “weekend”, and “peak” features was chosen based on these results to draw this impact.

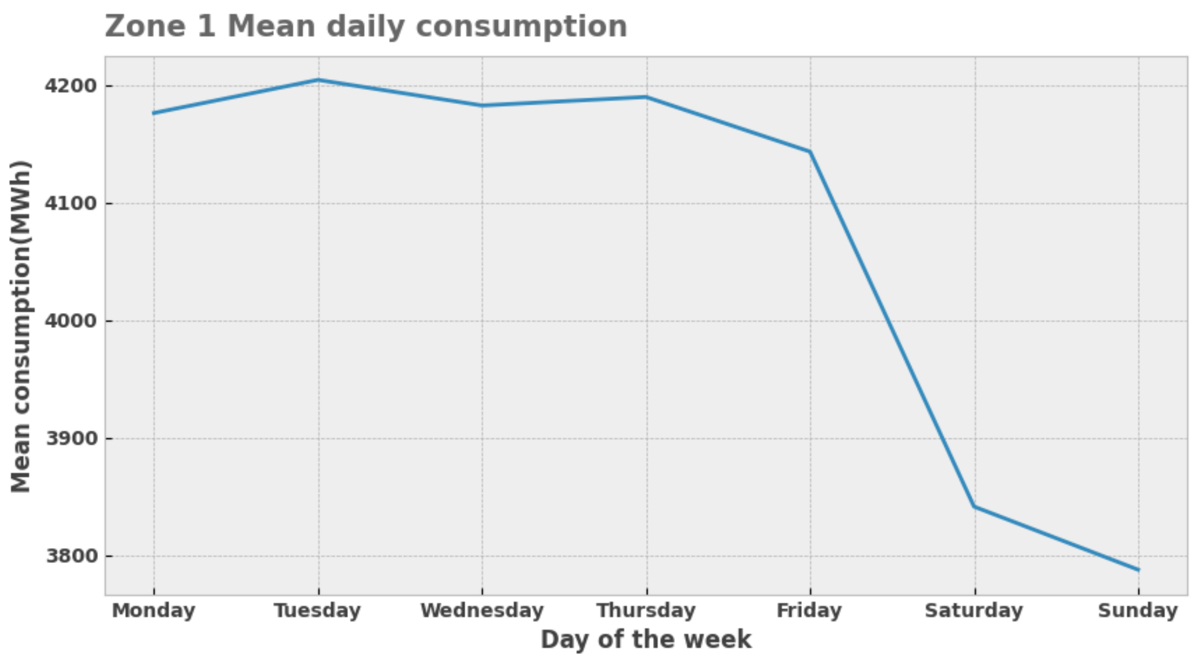


Fig 3 :A figure representing the mean daily electricity consumption in Norway.

Impact of Time

The utilization of electricity is significantly influenced by time, with energy consumption exhibiting fluctuations throughout the day, particularly during midday periods. To account for this temporal dependency, two functions are derived to represent the hour and day of the week. Additionally, external factors, such as seasonal variations, climate conditions, and holiday occurrences, can impact power load behavior. These factors, sourced from data outside the energy database, contribute to the overall variability in electricity usage.

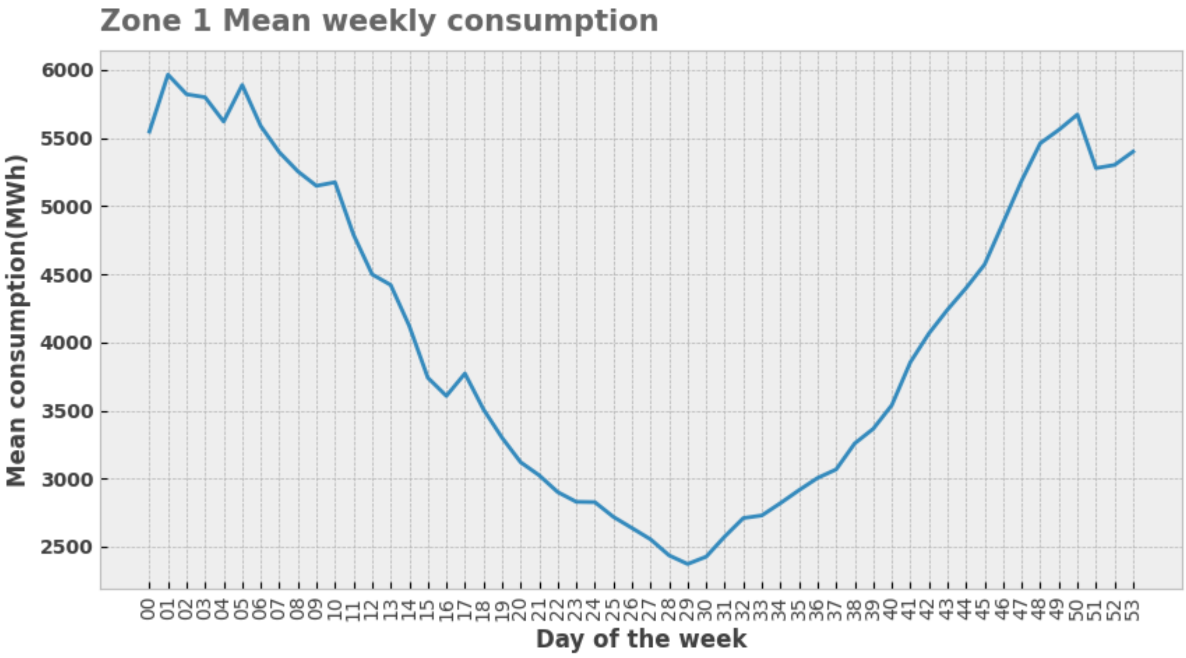


Fig 4: A figure representing the mean weekly electricity consumption in Norway.

From the figure above you can clearly tell the difference between between summer weeks and winter weeks.

In regions like Norway, electricity demand is heavily influenced by industrial activities from factories and offices. However, during weekends or holidays, when governmental and private offices are typically closed, industrial demand diminishes, leading to residential load dominance. Residential energy consumption is notably influenced by unpredictable human behaviors, contributing to lower forecasting accuracy during weekends and holidays. Conversely, weekdays exhibit more consistent patterns in electricity demand, both during daytime and evening hours. Furthermore, the demand during late-night and early-morning hours remains relatively stable across weekdays and weekends.

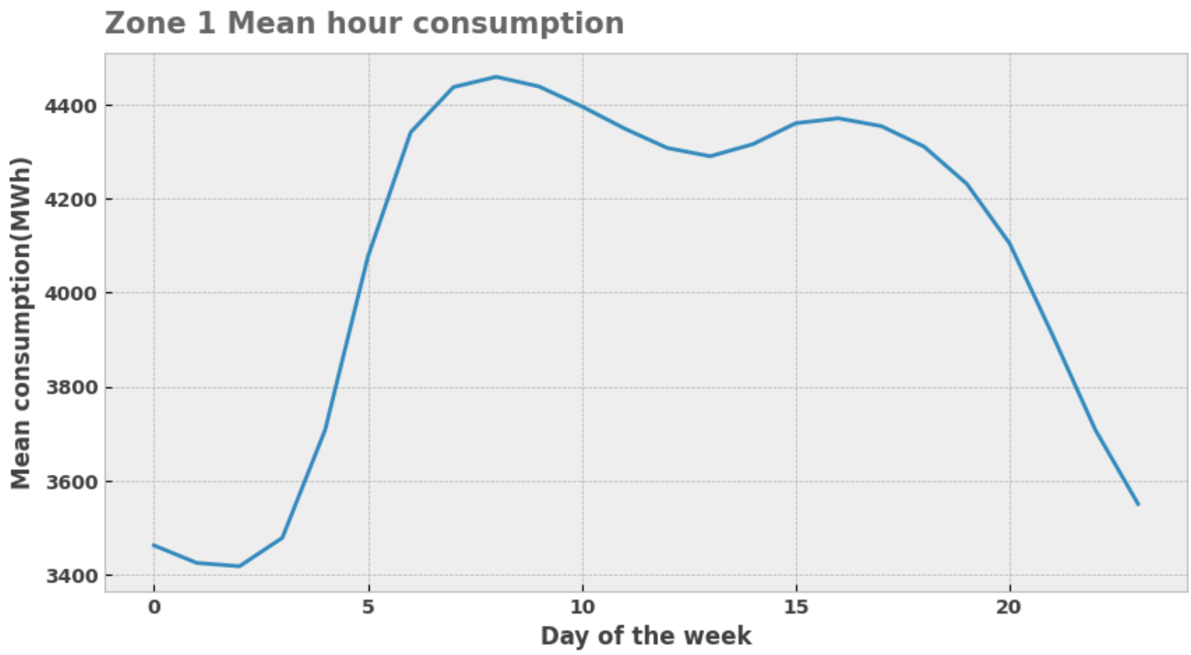


Fig 5:A figure representing the mean hourly electricity consumption in Norway.

Impact of Holidays

Norway's holiday season, marked by cultural festivities and winter traditions, presents a distinctive energy consumption landscape characterized by heightened electricity usage. Understanding the dynamics of holiday electricity consumption is imperative for optimizing energy infrastructure. Notably, the tradition of festive lighting displays emerged as a significant driver of energy demand, particularly in urban centers, where elaborate decorations adorn homes, streets, and public spaces.

Indoor heating systems experience a noticeable uptick in usage during the winter holidays, as households congregate for seasonal gatherings. Culinary customs, exemplified by the preparation of elaborate holiday meals utilizing electric appliances, also contribute to heightened energy consumption. Furthermore, outdoor activities and events, such as winter sports and festive markets, exert a discernible impact on electricity usage, necessitating additional lighting, heating, and infrastructure. The influx of tourists during the holiday season further amplifies energy demand in the hospitality sector, encompassing accommodations, dining establishments, and tourist attractions.

Some of the holidays include New year, Maundy Thursday, Good Friday, Easter Monday, Labour Day, Constitution Day, Ascension Day, Whit Monday, Christmas eve, Christmas day, Second Christmas.

Forecasting procedures

In this section, we'll dive into the detailed process of how we forecast electricity consumption. We'll explore the various steps and methods used to predict the amount of electricity that will be used over a given period.

Loading the dataset

A total of 81739 rows and 16 columns were loaded in for data analysis.

Datetime	load_no1	wind_no1	temp_no1	load_no2	wind_no2	temp_no2	load_no3	wind_no3	temp_no3	load_no4	wind_no4	temp_no4	load_no5	wind_no5	temp_no5
01/01/2015 00:00	4659.0	0.0	-2.5	4139.0	156.0	6.5	2370.0	259.0	3.0	2091.0	66.0	4.0	2211.0	19.0	6.5
01/01/2015 01:00	4552.0	0.0	-2.0	4039.0	159.0	7.0	2307.0	234.0	3.0	2078.0	69.0	4.0	2128.0	19.0	7.0
01/01/2015 02:00	4469.0	0.0	-1.5	3956.0	149.0	7.0	2273.0	200.0	3.0	2037.0	55.0	4.0	2148.0	19.0	7.0
01/01/2015 03:00	4442.0	0.0	0.0	3900.0	144.0	7.0	2286.0	192.0	2.0	2013.0	54.0	5.0	2114.0	19.0	7.0
01/01/2015 04:00	4488.0	0.0	0.0	3915.0	149.0	8.0	2333.0	192.0	2.0	2037.0	56.0	5.0	2131.0	19.0	8.0

Table 1: the first five rows of the dataset.

Columns

1. Datetime: This is the index in our time series that specifies the date and time associated with the consumption.
2. Load_no1: Electricity consumption for zones 1.
3. Load_no2: Electricity consumption for zones 2.
4. Load_no3: Electricity consumption for zones 3.
5. Load_no4: Electricity consumption for zones 4.
6. Load_no5: Electricity consumption for zones 5.

7. Temp_no1: Temperature consumption for zones 1.
8. Temp_no2: Temperature consumption for zones 2.
9. Temp_no3: Temperature consumption for zones 3.
10. Temp_no4: Temperature consumption for zones 4.
11. Temp_no5: Temperature consumption for zones 5.
12. Wind_no1: Windspeed consumption for zones 1.
13. Wind _no2: Windspeed consumption for zones 2.
14. Wind _no3: Windspeed consumption for zones 3.
15. Wind _no4: Windspeed consumption for zones 4.
16. Wind _no5: Windspeed consumption for zones 5.

Data preprocessing

We performed a preliminary descriptive statistical analysis of the datasets to address the issues of outliers.

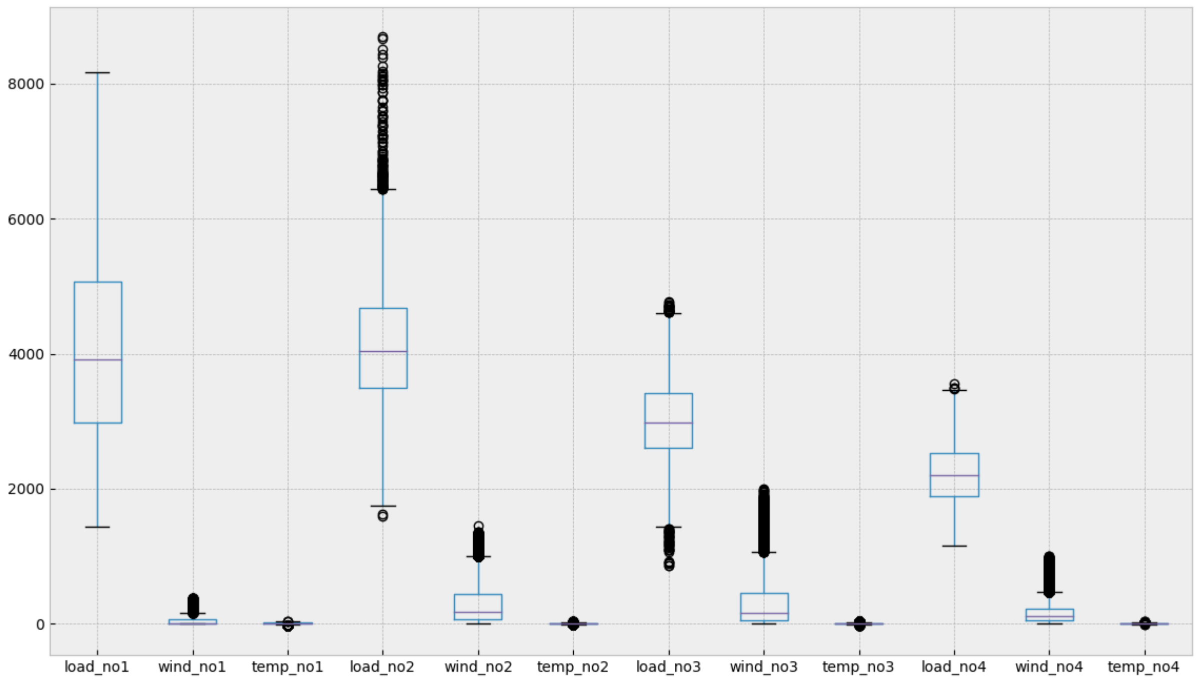


Fig6:A figure illustrating the outliers present in the data frame.

load_no1 has 15 outliers in total, which is 0.018% of data.

wind_no1 has 1867 outliers in total, which is 2.3% of data.

temp_no1 has 118 outliers in total, which is 0.14% of data.

load_no2 has 157 outliers in total, which is 0.19% of data.

wind_no2 has 484 outliers in total, which is 0.59% of data.

temp_no2 has 182 outliers in total, which is 0.22% of data.

load_no3 has 73 outliers in total, which is 0.089% of data.

wind_no3 has 1545 outliers in total, which is 1.9% of data.

temp_no3 has 260 outliers in total, which is 0.32% of data.

load_no4 has 11 outliers in total, which is 0.013% of data.

wind_no4 has 1647 outliers in total, which is 2.0% of data.

temp_no4 has 121 outliers in total, which is 0.15% of data.

load_no5 has 270 outliers in total, which is 0.33% of data.

wind_no5 has 3429 outliers in total, which is 4.2% of data.

temp_no5 has 182 outliers in total, which is 0.22% of data.

When processing the data, the scale and magnitude of the data may not be consistent. Thus, we also standardized the secondary data collected, following the Z-score standardization formula in the data standardization process. The previous data frame size was 81769 and the new data frame size became 73059.

Because of the above-mentioned factors that impact electricity consumption, more variables were created to improve the accuracy and reduce overfitting of the model. The variables added include month, day of the year, weekend, holiday, worktime, and peak.

To intuitively understand the distribution of data, we conducted a basic visual processing of the datasets. The distribution of electricity consumption across all five zones appears similar. However, Zone NO1 appears to have a wider distribution which may indicate that we have higher values of electricity consumption.

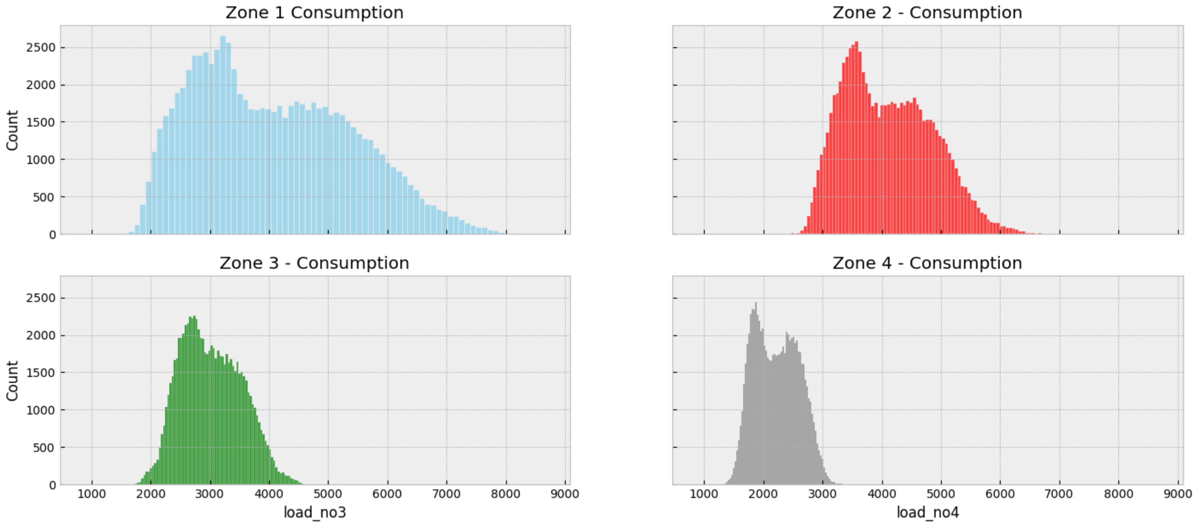


Fig 7:A figure representing the mean daily electricity consumption in Norway.

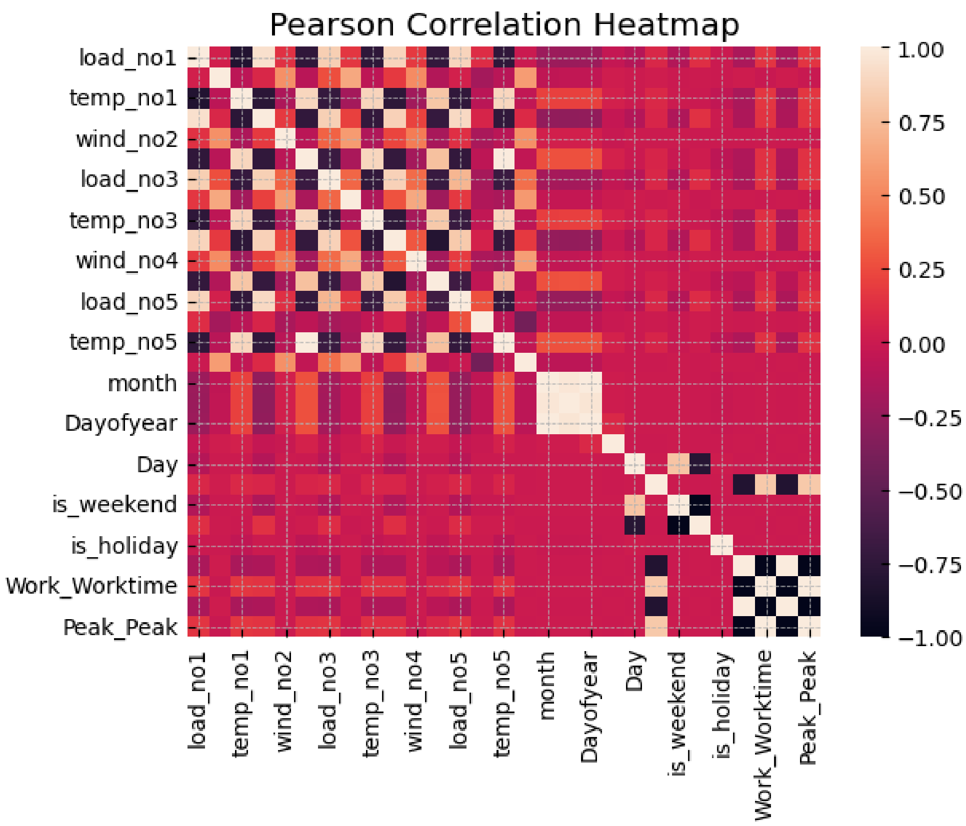


Fig 8:A Pearsons correlations Heat map illustrating the relationship between variables.

The study analyzed the correlation between variables using a heat map. Pearson's Correlation Coefficient helps to find out the relationship between two quantities. From the results below, some variables were positively correlated while others negative.

Training and tuning the electricity demand forecasting models.

For the univariate model like Arima, it is important to correct non-stationary data and remove seasonality. This will significantly impact the performance of the model.

1. Stationarity

A given time series is thought to consist of three systematic components including level, trend, seasonality, and one non-systematic component called noise. These components are defined as follows:

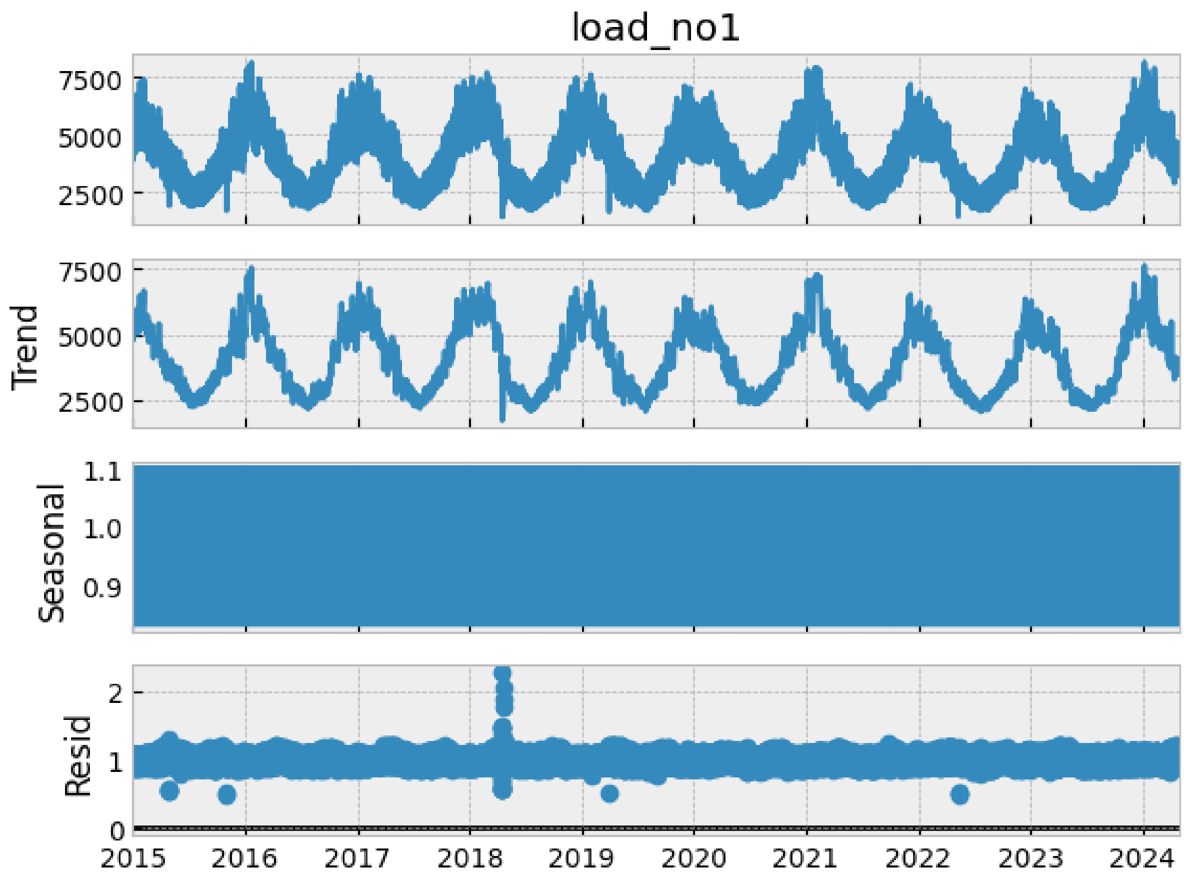
Level: The average value in the series.

Trend: The increasing or decreasing value in the series.

Seasonality: The repeating short-term cycle in the series.

Noise: The random variation in the series.

To perform a time series analysis, we may need to separate seasonality and trend from our series. The resultant series will become stationary through this process.



A figure representing trend, seasonality and residuals

There is an upward and downward trend and a recurring event where electricity consumption shoots maximum and minimum every year for all the zones. To test for stationarity, we must perform the ADF (Augmented Dickey-Fuller) Test.

1.1. ADF (Augmented Dickey-Fuller) Test

The Dickey-Fuller test is one of the most popular statistical tests. It can be used to determine the presence of unit root in the series, and hence help us understand if the series is stationary or not. The null and alternate hypothesis of this test is:

Null Hypothesis: The series has a unit root. (series is not stationary)

Alternate Hypothesis: The series has no unit root.

If we fail to reject the null hypothesis, we can say that the series is non-stationary. This means that the series can be linear or difference stationary (we will understand more about difference stationary in the next section). If both mean and standard deviation are flat lines (constant mean and constant variance), the series becomes stationary.

Series	ADF	P-Value	Num Of Lags	Num Of Observations	Critical Values
load_no1	-5.331974935747544	4.6995159683839465e-06	65	81703	{'1%': -3.430430039967484, '5%': -2.861575376324344, '10%': -2.5667888295
load_no2	-6.616094466873617	6.202030976603557e-09	65	81703	{'1%': -3.430430039967484, '5%': -2.861575376324344, '10%': -2.5667888295
load_no3	-6.052269565406708	1.2692467479622716e-07	65	81703	{'1%': -3.430430039967484, '5%': -2.861575376324344, '10%': -2.5667888295
load_no4	-6.500049520595319	1.167953335435998e-08	65	81703	{'1%': -3.430430039967484, '5%': -2.861575376324344, '10%': -2.5667888295
load_no5	-6.582042144424071	7.472199494498441e-09	65	81703	{'1%': -3.430430039967484, '5%': -2.861575376324344, '10%': -2.5667888295

In this case, the p-value is greater than 0.05 so we cannot reject the Null hypothesis. Also, the ADF test statistics are greater than the critical values. Therefore, the data is non-stationary. To get a stationary series, we need to eliminate the trend and seasonality from the series.

We started by taking a log of the series to reduce the magnitude of the values and reduce the rising trend in the series. Then after getting the log of the series, we find the rolling average of the series. A rolling average is calculated by taking input for the past 12 months and giving a mean consumption value at every point further ahead in series. After finding the mean, we take the difference of the series and the mean at every point in the series. This way, we eliminate trends out of a series and obtain a more stationary series.

Performed the Dickey-Fuller test (ADFT) once again. We performed this function to check whether the data is stationary or not.

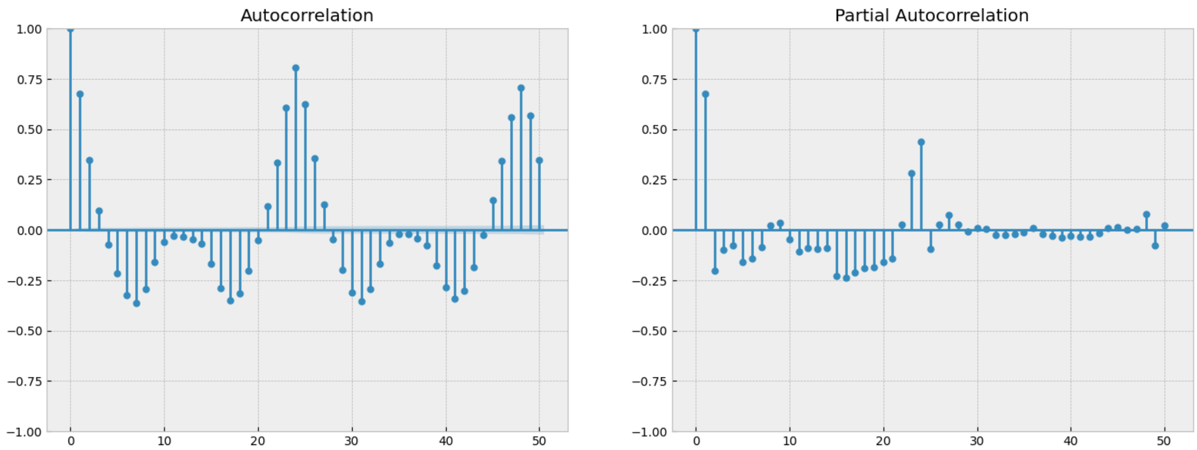
Series	ADF	P-Value	Num Of Lags	Num Of Observations	critical values
load_no1	-47.71983377684181	0.0	65	81692	{'1%': -3.430430050745373, '5%': -2.861575381087926, '10%': -2.5667888321302605}
load_no2	-48.87347024321383	0.0	65	81692	{'1%': -3.430430050745373, '5%': -2.861575381087926, '10%': -2.5667888321302605}
load_no3	-49.419137570948024	0.0	65	81692	{'1%': -3.430430050745373, '5%': -2.861575381087926, '10%': -2.5667888321302605}
load_no4	-48.78141447876949	0.0	65	81692	{'1%': -3.430430050745373, '5%': -2.861575381087926, '10%': -2.5667888321302605}
load_no5	-47.96702583673493	0.0	65	81692	{'1%': -3.430430050745373, '5%': -2.861575381087926, '10%': -2.5667888321302605}

From the above graph, we observed that the data attained stationarity. There can be cases when there is a high seasonality in the data. In those cases, just removing the trend will not help much. We also need to take care of the seasonality in the series. One such method for this task is differencing.

Differencing is a method of transforming a time series dataset. It can be used to remove the series dependence on time, so-called temporal dependence. This includes structures like trends and seasonality. It can help stabilize the mean of the time series by removing changes in the level of a time series, and so eliminating (or reducing) trend and seasonality. Differencing is performed by subtracting the previous observation from the current observation.

2. Seasonality

Before we go on to build our forecasting model, we need to determine optimal parameters for our model. For those optimal parameters, we need ACF and PACF plots.A nonseasonal ARIMA model is classified as an “ARIMA(p,d,q)” model, where: p → Number of autoregressive terms, d → Number of nonseasonal differences needed for stationarity, and q → Number of lagged forecast errors in the prediction equation. Values of p and q come through ACF and PACF plots.



A figure ACF AND PCF for Zone 1

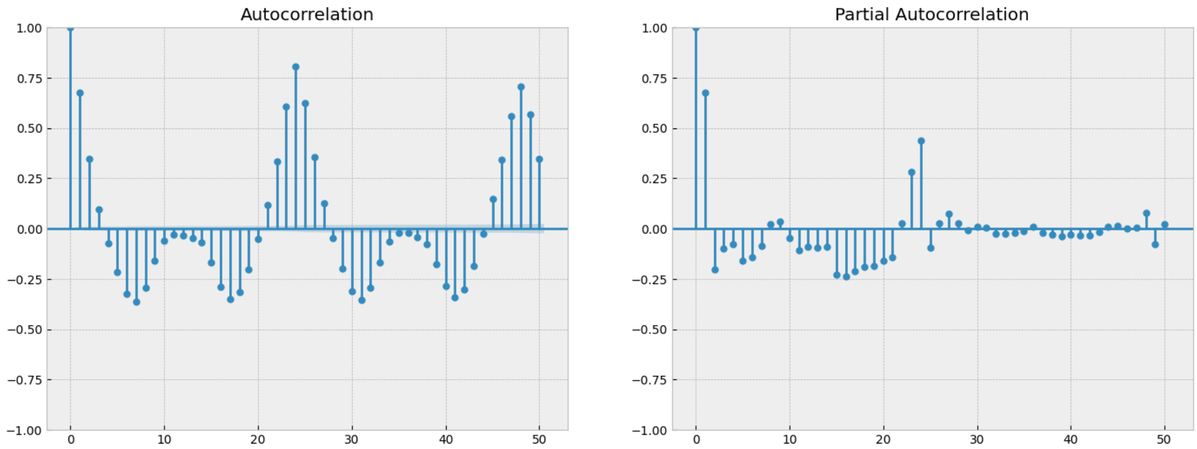
2.1.Autocorrelation and Partial Autocorrelation

Autocorrelation Function (ACF): Statistical correlation summarizes the strength of the relationship between two variables. Pearson’s correlation coefficient is a number between -1 and 1 that describes a negative or positive correlation respectively. A value of zero indicates no correlation.

We can calculate the correlation for time series observations with previous time steps, called lags. Because the correlation of the time series observations is calculated with values of the same series at previous times, this is called a serial correlation, or an autocorrelation. A plot of the autocorrelation of a time series by lag is called the Autocorrelation Function, or the acronym ACF. This plot is sometimes called a correlogram or an autocorrelation plot.

Partial Autocorrelation Function (PACF): A partial autocorrelation is a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed. The partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags.

The autocorrelation for observation and observation at a prior time step is comprised of both the direct correlation and indirect correlations. It is these indirect correlations that the partial autocorrelation function seeks to remove.



From the above plot we can derive the p, q values which are 3 and 2 respectively.

Forecasting Methods

Neural networks and machine learning techniques have gained traction in recent years due to their ability to handle complex data patterns, nonlinear relationships, and dynamic environments inherent in electricity consumption data. These methods offer a diverse range of algorithms and architectures tailored to different forecasting tasks, ranging from traditional approaches like linear regression to advanced deep learning models such as recurrent neural networks (Badar & Shams) and long short-term memory (LSTM) networks.

In this paper, we processed the collected time-series dataset to build the training data for predictive models. Following Waheed et al. [16], we used the linear regression model, support vector regression model, and random forest regression model for predicting the regional power demand. Regarding the support vector regression model, we used the grid search method for optimizing the model parameters. The values of the parameters of the SVR model and the random forest model are shown in Table 3 and Table 4, respectively. Note that a large power generation firm in Guangdong Province employs a simple year over-year method for creating the regional electricity demand forecast. The method examines the year-over-year changes from the previous month. By comparing the predicted values of the models with the actual values, we observed that the predictive ability of the selected models is satisfactory.

In this section we discuss the theory behind the statistical- and machine learning approaches is presented. The chapter ends with explaining the mathematics behind the models used in the thesis.

Method 1: Arima Model

Autoregressive Integrated Moving Average (ARIMA) models are a cornerstone in time series analysis, widely used for forecasting purposes. Understanding the theory behind both its statistical and machine learning aspects is crucial for effectively utilizing ARIMA models.

At its core, ARIMA combines autoregression (AR), differencing (I), and moving average (MA) components to model time series data. Let's break down each component:

> *Autoregression (AR):* Autoregression refers to modelling the relationship between an observation and several lagged observations. In ARIMA, the "AR" component captures the linear dependence between an observation and its previous observations, representing how past values influence future values.
>
> Mathematically, an AR(p) process can be expressed as:
>
> $yt=c+\phi 1yt-1+\phi 2yt-2+\ldots+\phi pyt-p+\epsilon tyt=c+\phi 1yt-1+\phi 2yt-2+\ldots+\phi pyt-p+\epsilon t$
>
> Here, $ytyt$ is the current observation, cc is a constant, $\phi 1$, $\phi 2,\ldots,\phi p\phi 1,\phi 2,\ldots,\phi p$ are the autoregressive parameters, $yt-1,yt-2,\ldots,yt-pyt-1,yt-2,\ldots,yt-p$ are the lagged observations, and $\epsilon t\epsilon t$ is white noise.
>
> *Differencing (I):* Differencing involves transforming a time series to make it stationary, i.e., with constant mean and variance over time. The differencing order, denoted by dd in ARIMA(d), indicates the number of times differencing is performed to achieve stationarity. This step removes trends and seasonality, making the series suitable for modelling.

Moving Average (MA): The moving average component represents the relationship between an observation and a residual error from a moving average model applied to lagged observations. In ARIMA, the "MA" component captures the influence of past forecast errors on the current observation.

An MA(q) process can be expressed as:

$y_t=c+\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\ldots+\theta_q\epsilon_{t-q}y_t=c+\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\ldots+\theta_q\epsilon_{t-q}$

Here, θ_1, $\theta_2,\ldots,\theta_q\theta_1,\theta_2,\ldots,\theta_q$ are the moving average parameters, and $\epsilon_{t-1},\epsilon_{t-2},\ldots,\epsilon_{t-q}\epsilon_{t-1},\epsilon_{t-2},\ldots,\epsilon_{t-q}$ are the residual errors from the lagged observations.

In the statistical framework, ARIMA models are estimated using maximum likelihood estimation (MLE) or least squares estimation. The parameters ($\phi\phi$'s, $\theta\theta$'s) are selected to minimize the model's residual sum of squares. Model diagnostics, such as residual analysis and goodness-of-fit tests, are crucial to assess the model's adequacy and identify potential issues like autocorrelation in residuals.

In recent years, machine learning techniques have been integrated with ARIMA models to improve forecasting accuracy. This includes using optimization algorithms like gradient descent to estimate parameters, as well as incorporating feature engineering and model selection techniques to enhance performance. Additionally, ensemble methods, such as combining multiple ARIMA models or integrating ARIMA with other forecasting algorithms, have been explored to capture complex patterns in time series data more effectively.

ARIMA models, while widely utilized for time series forecasting, exhibit several limitations that need to be acknowledged. Firstly, they assume a linear relationship between variables, which may not always hold true in real-world scenarios where nonlinear relationships are prevalent. Moreover, ARIMA models require the time series data to be stationary, a condition that may be challenging to achieve, particularly for datasets exhibiting non-stationary behavior. Additionally, while ARIMA can capture some forms of seasonality, it may struggle with datasets featuring complex seasonal patterns or multiple seasonality, necessitating alternative modeling approaches like SARIMA.

ARIMA models are generally not well-suited for long-term forecasting, as forecast accuracy tends to diminish with increasing forecast horizons. Sensitivity to outliers and anomalies in the data poses another challenge, potentially leading to distorted model estimation and poor forecasting performance. Moreover, selecting appropriate ARIMA model parameters can be complex and time-consuming, requiring iterative experimentation and model diagnostics. Data availability is also critical, as ARIMA models rely on a sufficient amount of historical data for accurate parameter estimation. Lastly, traditional ARIMA models do not incorporate exogenous variables, limiting their ability to account for external factors that may influence the time series. Despite these limitations, ARIMA models remain valuable tools for time series forecasting, with their shortcomings often mitigated through careful model selection and supplementary techniques.

Method 2: Linear Regression

Linear regression is a powerful statistical technique widely used for predicting continuous outcomes based on one or more predictor variables.

In the statistical framework, linear regression models are built on foundational principles of ordinary least squares (OLS) estimation. The objective is to find the line (or hyperplane in higher dimensions) that minimizes the sum of squared differences between the observed and predicted values. Mathematically, the linear regression model can be expressed as:

$Y=\beta 0+\beta 1X1+\beta 2X2+\ldots+\beta pXp+\epsilon Y=\beta 0+\beta 1X1+\beta 2X2+\ldots+\beta pXp+\epsilon$

Where:

- YY is the dependent variable (e.g., electricity consumption).
- $\beta 0\beta 0$ is the intercept term.
- $\beta 1,\beta 2,\ldots,\beta p\beta 1,\beta 2,\ldots,\beta p$ are the coefficients of the predictor variables ($X1,X2,\ldots,XpX1,X2,\ldots,Xp$).
- $X1,X2,\ldots,XpX1,X2,\ldots,Xp$ are the independent variables (e.g., temperature, time of day).
- $\epsilon\epsilon$ is the error term, representing the difference between the observed and predicted values.

The OLS estimation technique finds the values of $\beta 0$, $\beta 1,\ldots,\beta p\beta 0,\beta 1,\ldots,\beta p$ that minimize the sum of squared residuals. Model diagnostics, such as checking for multicollinearity, heteroscedasticity, and normality of residuals, are essential for assessing the model's adequacy and identifying potential issues.

Linear regression is viewed as a predictive modelling technique that learns from data to make accurate predictions. While the underlying principles of linear regression remain the same, machine learning introduces additional techniques to enhance model performance and flexibility.

Machine learning approaches often involve extensive feature engineering, where domain knowledge is used to create new predictor variables or transform existing ones to improve predictive accuracy. For example, features such as lagged values of electricity consumption or moving averages of temperature can be included to capture temporal patterns and seasonality.

Regularization techniques like Ridge regression and Lasso regression are commonly employed to prevent overfitting and improve the generalization ability of the model. These techniques introduce penalty terms on the size of the coefficients, effectively shrinking them towards zero and reducing model complexity.

Machine learning approaches typically employ cross-validation techniques to assess model performance on unseen data. K-fold cross-validation, for instance, partitions the dataset into k subsets, trains the model on k-1 subsets, and evaluates its performance on the remaining subset. This helps to obtain more reliable estimates of the model's predictive performance.

Method 3: XGBoost Regressor

XGBoost, short for Extreme Gradient Boosting, stands out as a powerful and versatile algorithm in the realm of machine learning, particularly renowned for regression tasks. Its popularity stems from its ability to construct an ensemble of decision trees, effectively leveraging their collective predictive power to minimize errors between predicted and actual values.

XGBoost can be expressed as follows:

$$\hat{y}i = \sum k = 1K f k(x_i),$$

where:

- $\hat{y}i$ represents the predicted value for the *i*-th observation.
- K denotes the total number of trees in the ensemble.
- $fk(xi)$ represents the prediction of the k-th tree for the i-th observation xi.

Each individual tree prediction $fk(xi)$ can be further decomposed into a sum of terminal region predictions:

$$fk(xi) = wq(xi),$$

where:

- $q(xi)$ represents the index of the terminal region to which $xixi$ belongs.
- ww denotes the output value associated with the terminal region.

In XGBoost, the predicted value $\hat{y}i$ is obtained by summing the predictions of all trees in the ensemble.

The objective of XGBoost is to minimize a loss function LL that measures the discrepancy between the predicted values and the true values:

$$Obj = \sum i = 1nL(yi, \hat{y}i) + \sum k = 1K\Omega(fk)$$

where:

- n denotes the total number of observations.
- yi represents the true value for the ii-th observation.
- $\Omega(f_k)$is a regularization term that penalizes complex models to prevent overfitting.

The model parameters, including the structure of each tree and the weights associated with terminal regions, are optimized through gradient boosting, which iteratively updates the model to minimize the objective function.

At its essence, XGBoost operates by sequentially building decision trees, with each subsequent tree correcting the errors made by the ensemble of preceding trees. This iterative process allows XGBoost to gradually improve its predictive performance, ultimately producing a robust and accurate regression model. Importantly, XGBoost incorporates regularization techniques to prevent overfitting, ensuring that the model generalizes well to unseen data. By penalizing overly complex models, regularization helps strike a balance between bias and variance, leading to more reliable predictions.

One key feature of XGBoost is its ability to provide insights into the importance of different features in the dataset. By analysing the contribution of each feature to the model's predictive performance, practitioners can gain valuable insights into the underlying relationships driving the phenomenon being modelled. This feature importance analysis enables stakeholders to prioritize factors that have the most significant impact on the outcome variable, guiding decision-making processes and informing resource allocation strategies.

The regression process with XGBoost typically follows a systematic series of steps. It begins with splitting the dataset into training and testing sets, ensuring that the model's performance is evaluated on unseen data. Next, the model is initialized with default parameters, setting the stage for subsequent training iterations. During the training phase, XGBoost iteratively builds decision trees, optimizing them to minimize a predefined loss function, such as mean squared error (MSE) for regression tasks. The model's performance is continually evaluated on the validation set, allowing for early stopping to prevent overfitting.

Hyperparameter tuning is a critical aspect of the XGBoost regression process, aiming to optimize model performance by selecting the most effective combination of hyperparameters. Techniques such as grid search or random search can be employed to systematically explore the hyperparameter space and identify the configuration that yields the best results. This iterative process involves adjusting parameters such as the learning rate, tree depth, and regularization parameters, fine-tuning the model to achieve optimal performance.

Once the model has been trained and tuned, it can be deployed to make predictions for new, unseen data. Leveraging the insights gained from feature importance analysis, stakeholders can interpret the model's predictions and make informed decisions based on the underlying patterns uncovered by XGBoost.

Method 4: Ridge regression

Ridge regression, also known as Tikhonov regularization or L2 regularization, is a linear regression technique that introduces a penalty term to the standard least squares method. This penalty term helps mitigate overfitting by penalizing large coefficients, encouraging simpler models with smaller parameter values. Ridge regression is particularly useful when dealing with multicollinearity, where predictor variables are highly correlated with each other.

The equation for ridge regression can be expressed as:

$$\hat{y} = \beta 0 + \beta 1x1 + \beta 2x2 + \cdots + \beta pxp + \lambda \sum j = 1p\beta j2,$$

where:

- ŷ represents the predicted value.
- $\beta 0$, $\beta 1$,…,βp are the regression coefficients.
- $x1$,$x2$,…,xp are the predictor variables.
- λ is the regularization parameter, also known as the tuning parameter, which controls the strength of the penalty term.

The objective of ridge regression is to minimize the following cost function:

$$\text{Cost} = \sum i = 1n(yi - y\hat{}i)2 + \lambda \sum j = 1p\beta j2,$$

where:

- n is the number of observations.
- yi is the true value for the ii-th observation.
- $\hat{y}i$ is the predicted value for the ii-th observation.

The first term in the cost function represents the residual sum of squares, which measures the discrepancy between the predicted and true values. The second term is the penalty term, which penalizes large coefficient values. The regularization parameter $\lambda\lambda$ controls the trade-off between minimizing the residual sum of squares and minimizing the size of the coefficients.

By adding the penalty term to the cost function, ridge regression shrinks the coefficient estimates towards zero, reducing their variance and mitigating multicollinearity issues. This helps improve the model's generalization ability and reduces the risk of overfitting, particularly when dealing with high-dimensional datasets or when predictor variables are highly correlated.

In practice, the value of the regularization parameter λ is chosen through techniques such as cross-validation, where different values of λ are evaluated, and the one that yields the best model performance is selected. Ridge regression is widely used in various fields, including finance, economics, and engineering, where predictive modeling is applied to complex datasets with correlated predictors.

Method 5: MLP

A Multi-Layer Perceptron (MLP) is a type of artificial neural network (ANN) that consists of multiple layers of nodes, or neurons, arranged in a feedforward manner. MLPs are versatile models capable of approximating complex nonlinear relationships between input and output variables. They are commonly used for both regression and classification tasks in machine learning.

The basic architecture of an MLP consists of three types of layers:

1. Input Layer: This layer consists of neurons that receive the input data. Each neuron corresponds to a feature or attribute of the input data.
2. Hidden Layers: These layers are composed of one or more layers of neurons situated between the input and output layers. The neurons in the hidden layers perform computations on the input data, transforming it into a form that can be used to make predictions.
3. Output Layer: This layer consists of neurons that produce the model's output. The number of neurons in the output layer depends on the nature of the prediction task. For regression tasks, there is typically a single neuron producing a continuous output value, while for classification tasks, there may be multiple neurons, each corresponding to a class label.

The connections between neurons in adjacent layers are characterized by weights, which represent the strength of the connections. During the training process, these weights are adjusted iteratively using optimization algorithms such as gradient descent to minimize a loss function, which measures the discrepancy between the predicted and true values.

MLPs are capable of learning complex patterns in the data through a process known as backpropagation. In backpropagation, the error from the output layer is propagated backward through the network, and the weights are updated accordingly to reduce the error. This process is repeated iteratively until the model converges to a set of weights that minimize the loss function.

One of the key advantages of MLPs is their ability to learn nonlinear relationships in the data, making them suitable for a wide range of tasks across various domains. However, MLPs are also prone to overfitting, particularly when the model architecture is complex or when the training data is limited. Regularization techniques such as dropout and weight decay can be used to mitigate overfitting and improve generalization performance.

In summary, MLPs are powerful and flexible models capable of learning complex patterns in data. By leveraging multiple layers of neurons and sophisticated optimization algorithms, MLPs can achieve state-of-the-art performance in a wide range of machine learning tasks. However, careful tuning of hyperparameters and regularization techniques is often necessary to ensure optimal performance and prevent overfitting.

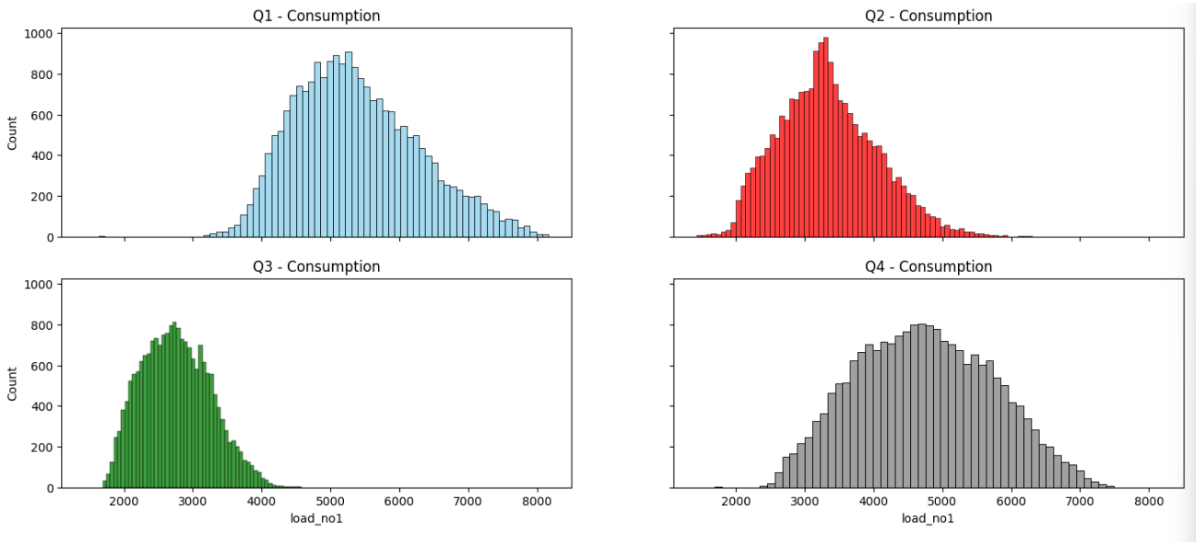
Chapter 5: Results/discussion.

This chapter outlines the findings of our analysis and models. We will briefly explain the facts and results from our models and present illustrations to visualize these results using graphs and tables. After the results are presented, interpretation of the results will be discussed concerning the objective of the thesis.

Brief overview of bidding zones in the country studied.

Eastern Norway (Zone 1)

Eastern Norway is the most populous and economically significant region in the country, home to cities like Oslo, Drammen, and Fredrikstad. The climate in Eastern Norway varies from coastal areas to inland regions, but overall, it experiences colder winters and warmer summers compared to Southern Norway.



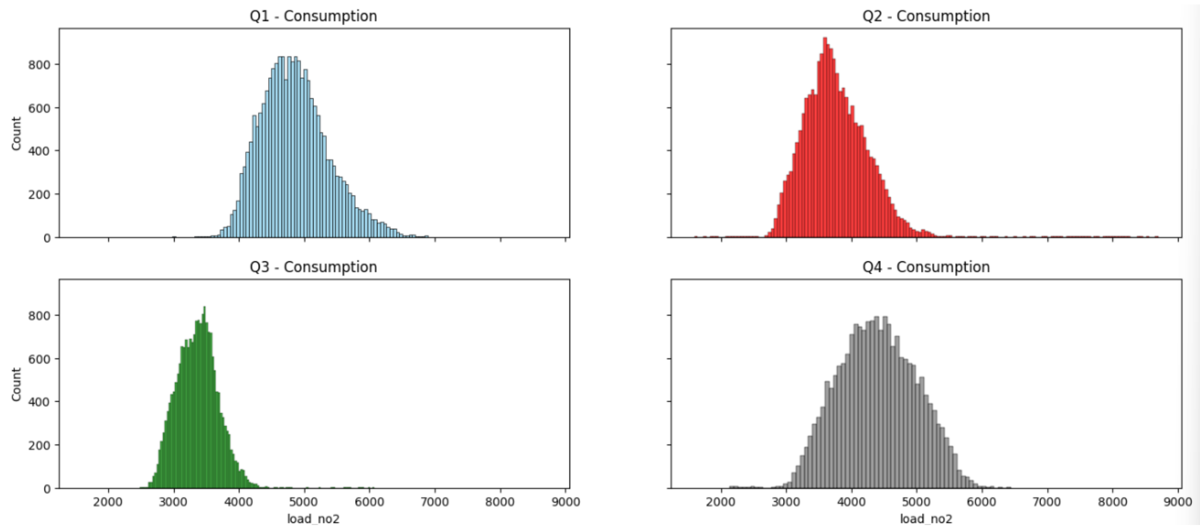
The winter season in Eastern Norway sees a significant increase in electricity consumption due to heating requirements. Residential, commercial, and industrial buildings rely on electric heating systems to maintain comfortable indoor temperatures, contributing to higher energy demand.

Holidays such as Christmas and Easter amplify electricity consumption in Eastern Norway. Decorative lighting, both indoors and outdoors, illuminates homes, streets, and public spaces, creating a festive atmosphere. Traditional holiday meals are prepared using electric appliances, and gatherings with family and friends further contribute to the overall rise in energy usage.

During the summer months, electricity consumption in Eastern Norway experiences a slight decrease compared to winter. Warmer temperatures reduce the need for electric heating, although increased use of electric cooling systems and appliances may partially offset this decrease.

Southern Norway (Zone 2)

Southern Norway encompasses regions such as Oslo, the capital city, and the surrounding areas. The climate in Southern Norway is relatively mild compared to the northern parts of the country, with warmer temperatures and less severe winters. Despite this, electricity consumption in Southern Norway experiences noticeable fluctuations throughout the year.



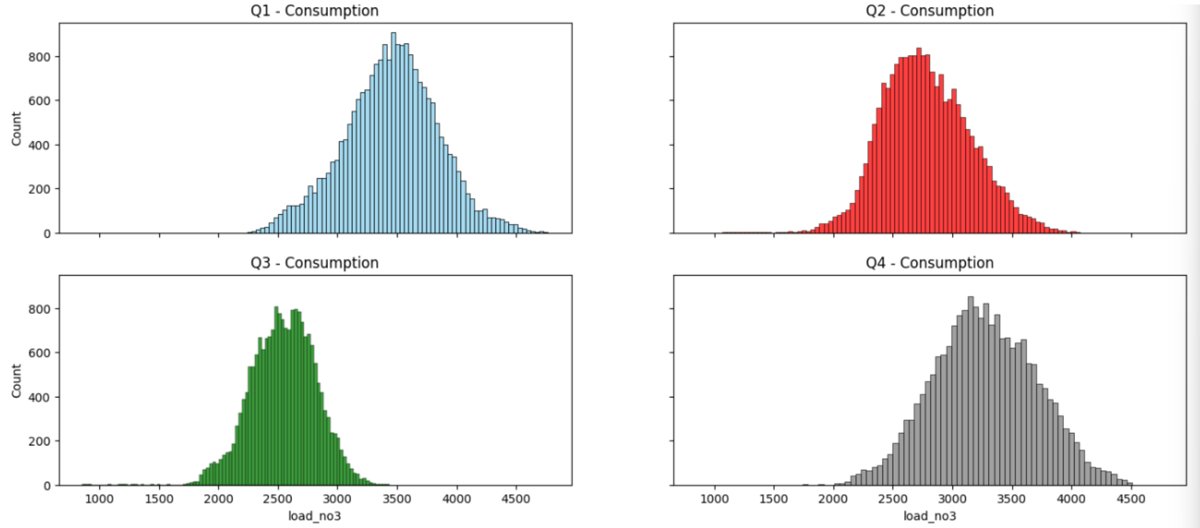
During the winter months, which typically span from November to March, electricity consumption in Southern Norway peaks. This increase can be attributed to the widespread use of electric heating systems in homes, offices, and other buildings. Additionally, the shorter daylight hours during winter necessitate more artificial lighting, further contributing to the rise in electricity usage.

Holidays such as Christmas and New Year's Eve also significantly impact electricity consumption in Southern Norway. Decorative lighting, both indoors and outdoors, adds to the festive atmosphere but also increases energy demand. Cooking large meals and hosting gatherings further elevate electricity usage during these holidays.

In contrast, during the summer months, electricity consumption in Southern Norway tends to decrease. Warmer temperatures reduce the need for electric heating, and longer daylight hours result in less reliance on artificial lighting. However, increased use of electric cooling systems, such as air conditioners and fans, during heatwaves can partially offset this decrease in consumption.

Central Norway (Zone 3)

Central Norway is characterized by its diverse landscapes, including mountains, valleys, and coastal areas. The region experiences colder winters and milder summers compared to Southern Norway. Electricity consumption patterns in Central Norway are influenced by factors such as geography, climate, and economic activities.



Winter in Central Norway brings colder temperatures and snowfall, leading to higher electricity usage for heating purposes. Mountainous areas attract tourists for skiing and snowboarding activities, further impacting energy consumption in hotels, resorts, and recreational facilities.

Holidays like Christmas and Easter see a notable increase in electricity consumption in Central Norway as well. Festive lighting, decorations, and indoor activities contribute to the overall rise in energy demand during these periods.

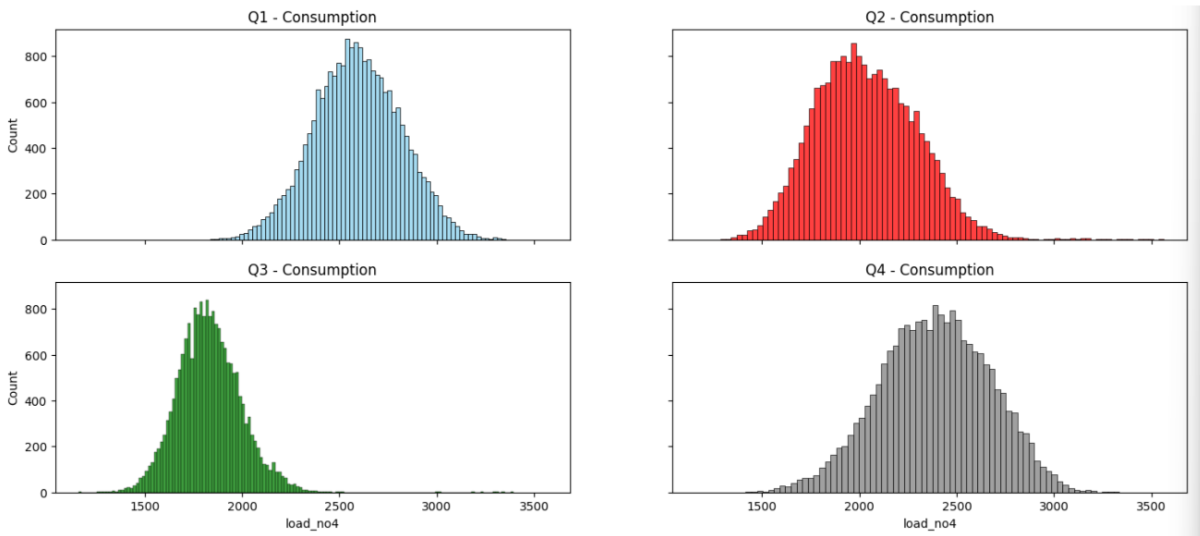
During the summer, electricity consumption in Central Norway remains relatively stable. While warmer temperatures may lead to increased use of electric cooling systems in some areas, particularly coastal regions, the overall impact on electricity demand is moderate compared to other seasons.

Northern Norway (Zone 4)

Northern Norway is characterized by its Arctic climate, with long, dark winters and extended daylight hours in summer. The region experiences the coldest temperatures in the country, leading to high electricity consumption for heating purposes during the winter months.

Winter in Northern Norway sees a significant increase in electricity consumption as residents rely heavily on electric heating systems to combat the cold. Additionally, shorter daylight hours necessitate more artificial lighting, further contributing to the rise in energy demand.

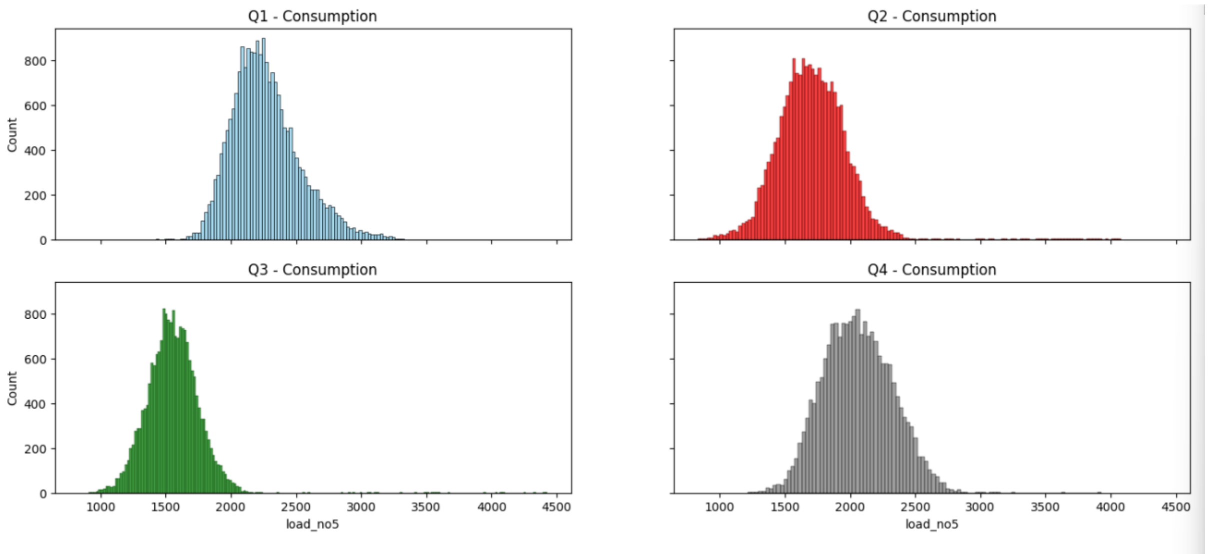
Holidays like Christmas and New Year's Eve witness a surge in electricity consumption in Northern Norway as well. Festive lighting and decorations brighten up homes and public spaces, adding to the overall increase in energy usage. Traditional holiday meals and gatherings with family and friends also contribute to higher electricity demand during these periods.



During the summer, electricity consumption in Northern Norway experiences a decrease compared to winter. Warmer temperatures reduce the need for electric heating, although increased use of electric cooling systems and appliances may partially offset this decrease.

Western Norway (Zone 5)

Western Norway is characterized by its stunning fjords, mountains, and coastal landscapes. The region experiences a maritime climate with ample rainfall throughout the year. Electricity consumption patterns in Western Norway are influenced by both natural and human factors.



Winter in Western Norway brings colder temperatures, particularly in higher elevations, leading to higher electricity usage for heating purposes. The rugged terrain and scattered population centres present challenges for infrastructure maintenance, which can impact electricity distribution and reliability during extreme weather events.

Holidays like Christmas and Easter see a surge in electricity consumption in Western Norway as well. Festive lighting and decorations adorn homes, streets, and public spaces, contributing to the overall increase in energy demand. Additionally, indoor activities such as cooking traditional holiday meals and hosting gatherings further elevate electricity usage.

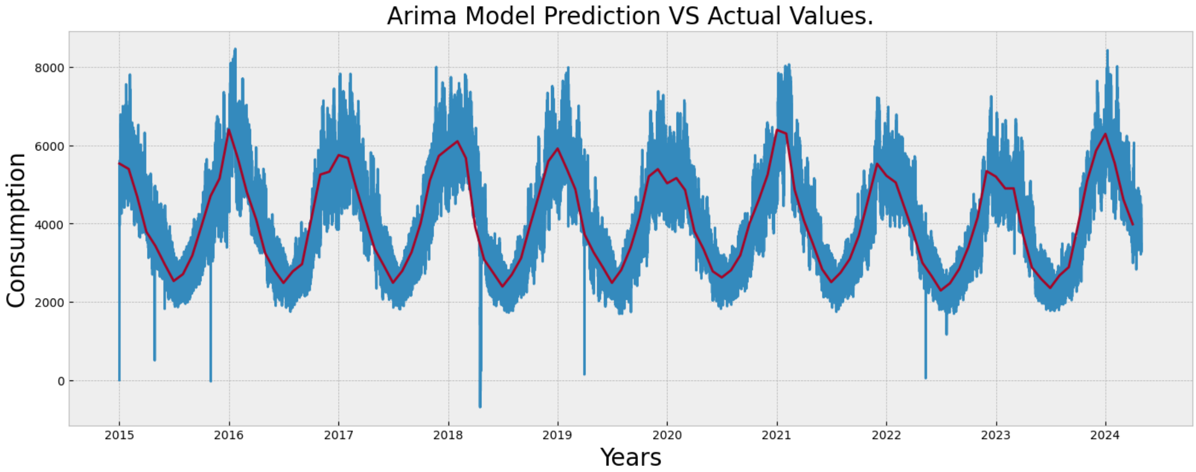
During the summer, electricity consumption in Western Norway remains relatively stable. While warmer temperatures may lead to increased use of electric cooling systems, particularly in coastal areas, the overall impact on electricity demand is less pronounced compared to other regions with hotter climates.

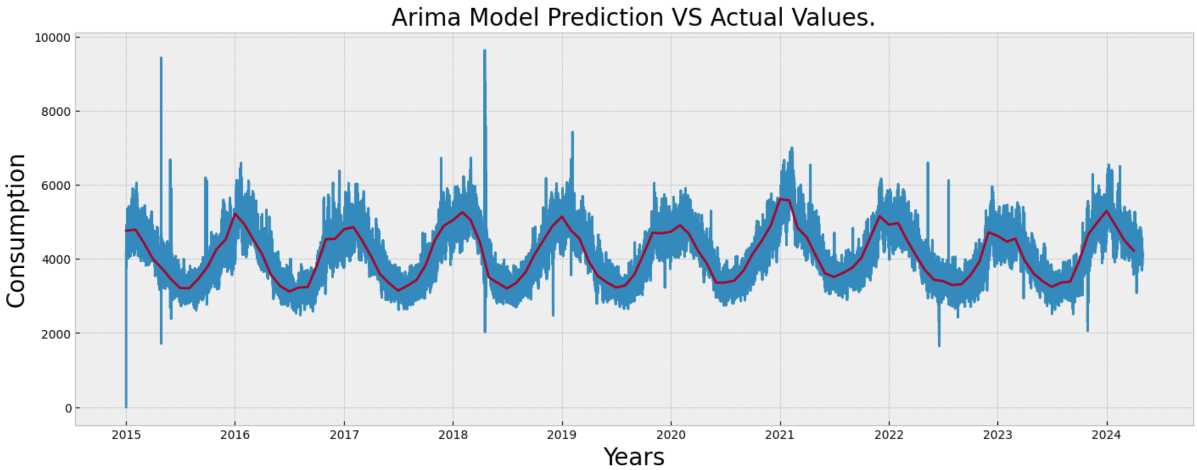
The seasonal electricity consumption patterns of the five different Norwegian zones vary depending on factors such as climate, geography, cultural traditions, and economic activities. Understanding these patterns is crucial for energy management, infrastructure planning, and policy development to ensure reliable and sustainable energy supply across the country.

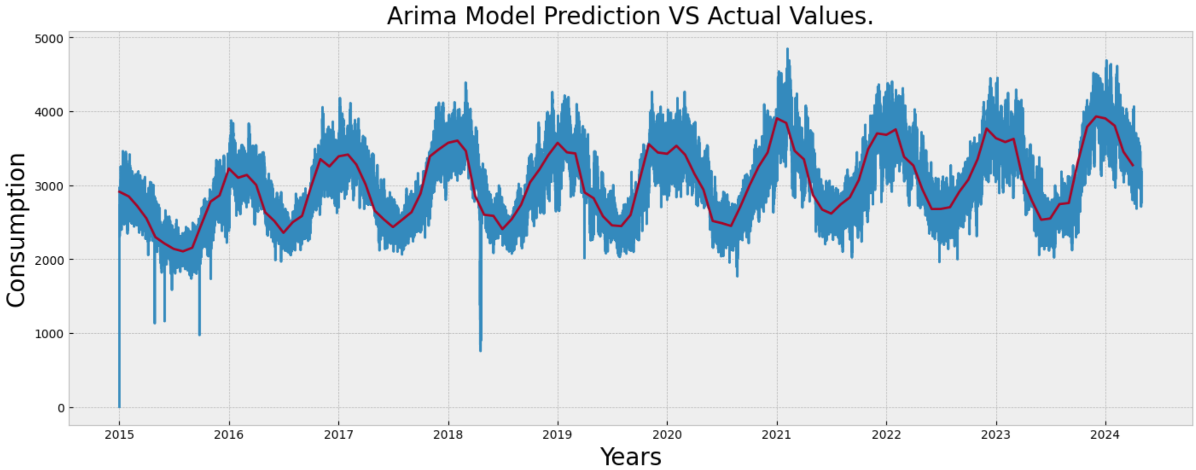
ARIMA Model Performance

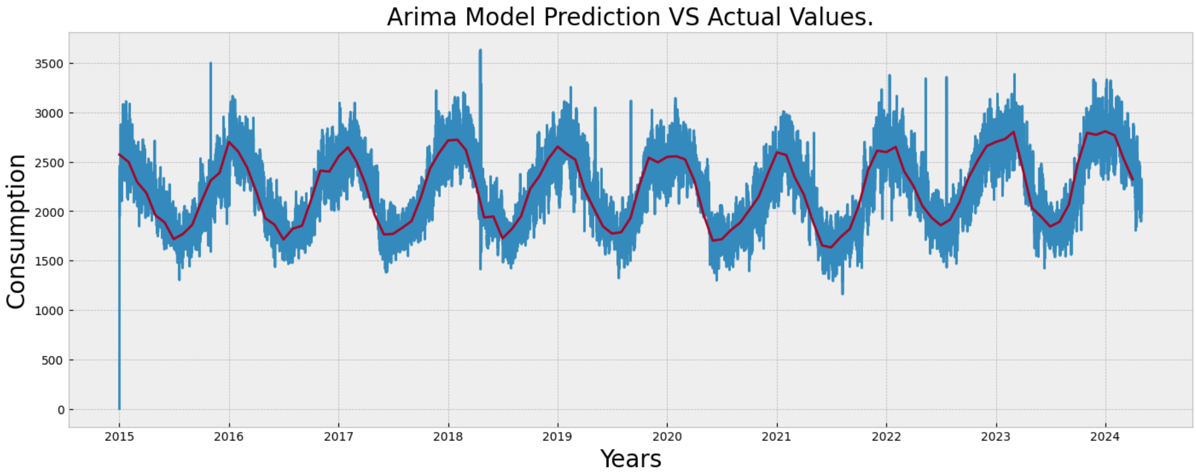
In this study, we assessed the performance of the ARIMA model in forecasting electricity consumption across five distinct geographical zones. Our analysis aimed to evaluate the model's ability to accurately predict electricity usage patterns, thereby providing valuable insights for energy management and decision-making processes.

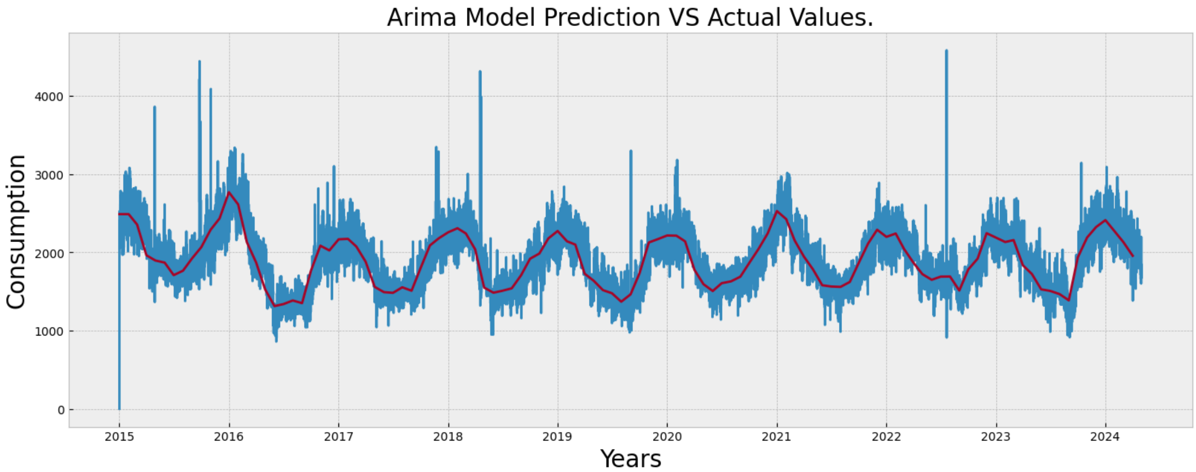
Upon rigorous evaluation, the ARIMA model demonstrated exceptional predictive accuracy across all zones, as evidenced by the close alignment between the actual and predicted values of electricity consumption.











The success of the ARIMA model can be attributed to its inherent flexibility and adaptability to diverse data patterns. By leveraging autoregression, differencing, and moving average components, the model effectively captured the underlying temporal dependencies present in the electricity consumption data. Furthermore, meticulous parameter tuning, and model optimization techniques were employed to enhance the model's predictive performance and generalization capabilities.

The close alignment between observed and predicted values across all zones underscores the robustness and reliability of the ARIMA model in forecasting electricity consumption. These findings hold significant implications for energy planners, policymakers, and stakeholders involved in energy management initiatives. By providing accurate and reliable predictions, the ARIMA model facilitates informed decision-making processes, enabling proactive measures to optimize energy utilization and enhance sustainability.

Despite the promising results obtained, it is important to acknowledge certain limitations inherent in the study. Factors such as data quality, model assumptions, and external influences may have influenced the performance of the ARIMA model and warrant further investigation in future research endeavors. Additionally, ongoing efforts to refine and enhance the model's performance through advanced modeling techniques and data integration strategies are essential for addressing evolving challenges in energy forecasting and management.

In conclusion, the findings of this study highlight the efficacy of the ARIMA model as a valuable tool for forecasting electricity consumption across diverse geographical zones. By leveraging its predictive capabilities, the ARIMA model contributes to advancing the field of energy forecasting and sustainability, ultimately supporting efforts towards efficient energy utilization and environmental conservation.

Forecasting the Results by zone

Zone 1

Model Performance

XGBoost:
MSE: 19949.31
R2: 0.99

Random Forest:
MSE: 19715.52
R2: 0.99

Linear Regression:
MSE: 449408.64
R2: 0.75

MLP:
MSE: 99441.68
R2: 0.94

Ridge Regression:
MSE: 449471.39
R2: 0.75

Zone 2

Model Performance

XGBoost:	Random Forest:	Linear Regression:
MSE: 19849.52	MSE: 19782.49	MSE: 449390.12
R2: 0.98	R2: 0.98	R2: 0.73
MLP:	Ridge Regression:	
MSE: 99464.82	MSE: 449525.18	
R2: 0.94	R2: 0.76	

Zone 3

Model Performance

XGBoost:	Random Forest:	Linear Regression:
MSE: 20012.83	MSE: 19825.91	MSE: 450112.47
R2: 0.97	R2: 0.97	R2: 0.71
MLP:	Ridge Regression:	
MSE: 99761.29	MSE: 450231.84	
R2: 0.93	R2: 0.70	

Zone 4

Model Performance

XGBoost:	Random Forest:	Linear Regression:
MSE: 20158.47	MSE: 19936.25	MSE: 450567.82
R2: 0.97	R2: 0.97	R2: 0.70
MLP:	Ridge Regression:	

MSE: 100086.29	MSE: 450678.91
R2: 0.93	R2: 0.69

Zone 5

Model Performance

XGBoost:	Random Forest:	Linear Regression:
MSE: 19785.62	MSE: 19696.84	MSE: 449950.13
R2: 0.98	R2: 0.98	R2: 0.71

MLP:	Ridge Regression:
MSE: 99357.74	MSE: 450089.25
R2: 0.94	R2: 0.70

Analysis and Discussion

Zone 1

The provided model performance metrics offer a comprehensive overview of how various machine learning algorithms perform in predicting electricity consumption. These metrics serve as valuable benchmarks for evaluating the efficacy of each algorithm, shedding light on their strengths and weaknesses in capturing the underlying patterns in the data.

Starting with XGBoost and Random Forest, both ensemble learning methods exhibit notably low mean squared error (MSE) values of 19949.31 and 19715.52, respectively. This suggests that they excel in minimizing the squared differences between predicted and actual values, indicating their robust predictive capabilities. Moreover, achieving high $R2$ scores of 0.99, these algorithms can explain approximately 99% of the variance in the electricity consumption data, highlighting their exceptional accuracy and ability to capture complex relationships within the dataset. Ensemble methods like XGBoost and Random Forest leverage the collective predictive power of multiple decision trees, enabling them to model intricate nonlinear relationships effectively.

In contrast, the linear regression models—Linear Regression and Ridge Regression—yield substantially higher MSE values of 449408.64 and 449471.39, respectively. These results indicate comparatively poorer predictive performance when compared to ensemble methods. The lower $R2R2$ scores of 0.75 suggest that linear regression models can only explain about 75% of the variance in the data. This discrepancy may stem from the inherent assumption of linearity in these models, which may not adequately capture the complex nonlinear patterns present in electricity consumption data. Despite this, linear regression-based approaches still offer valuable insights into the predictive power of simpler models and can be useful in scenarios where interpretability is paramount.

The MLP model, while exhibiting higher MSE compared to ensemble methods, demonstrates a relatively strong performance with an MSE of 99441.68. Its $R2$ score of 0.94 indicates that it can explain approximately 94% of the variance in the data, highlighting its effectiveness in capturing underlying patterns in electricity consumption. MLPs, being versatile and capable of learning complex nonlinear relationships, offer a middle ground between the simplicity of linear models and the complexity of ensemble methods.

Zone 2

XGBoost and Random Forest, both ensemble learning techniques demonstrate impressive prowess with minimal MSE values of 19849.52 and 19782.49, correspondingly. These values imply that both algorithms adeptly minimize the discrepancies between predicted and actual values, indicative of robust predictive capabilities. Furthermore, achieving lofty $R2$ scores of 0.98, these models proficiently capture about 98% of the variability in the electricity consumption data, underscoring their reliability and accuracy.

As for the linear regression models, namely Linear Regression and Ridge Regression, we observe relatively elevated MSE values compared to ensemble methods. Linear Regression yields an MSE of 449390.12, while Ridge Regression exhibits a slightly lower MSE of 449525.18. Despite these higher MSE values, both models attain respectable $R2$ scores of 0.73 and 0.76, respectively. These scores suggest that linear regression-based approaches can elucidate approximately 73% to 76% of the variance in the data. While these models may not rival ensemble methods in predictive accuracy, they offer simplicity and ease of interpretation, rendering them valuable for comprehending the relationships between predictor variables and electricity consumption.

The MLP model, while manifesting higher MSE compared to ensemble methods, still showcases commendable performance with an MSE of 99464.82. Its $R2$ score of 0.94 indicates that it can elucidate about 94% of the variance in the data, underscoring its adeptness in capturing underlying patterns in electricity consumption. MLPs, being versatile and proficient in discerning complex nonlinear relationships, offer a middle ground between intricacy and interpretability. However, the higher MSE compared to ensemble methods suggests the potential for further refinement in the model's predictive accuracy.

While ensemble methods like XGBoost and Random Forest excel in accuracy and resilience, linear regression-based approaches provide simplicity and interpretability. The MLP model bridges the divide between complexity and interpretability, offering a flexible framework for discerning intricate patterns in the data. Understanding the trade-offs and strengths of each algorithm is imperative for selecting the most suitable model for electricity consumption forecasting, ensuring optimal performance and well-informed decision-making in energy management.

Zone 3

The provided model performance metrics offer a detailed evaluation of the predictive capabilities of various machine learning algorithms applied to the task of electricity consumption forecasting. These metrics, consisting of Mean Squared Error (MSE) and R-squared ($R2$), serve as quantitative benchmarks for assessing the accuracy and explanatory power of each model.

Commencing with ensemble methods, XGBoost and Random Forest, both exhibit competitive performance with relatively low MSE values of 20012.83 and 19825.91, respectively. These values indicate the algorithms' effectiveness in minimizing the discrepancies between predicted and observed values. Additionally, achieving high $R2$ scores of 0.97 for both models underscore their capacity to explain approximately 97% of the variance in electricity consumption data, reflecting robust predictive capabilities and a strong ability to capture underlying patterns in the dataset.

Transitioning to linear regression-based approaches, Linear Regression and Ridge Regression, we observe higher MSE values compared to ensemble methods. Linear Regression yields an MSE of 450112.47, while Ridge Regression exhibits a slightly lower MSE of 450231.84.

Despite these elevated MSE values, both models achieve respectable $R2$ scores of 0.71 and 0.70, respectively. This suggests that linear regression-based models can elucidate approximately 71% to 70% of the variance in the data. While they may not match the predictive accuracy of ensemble methods, their simplicity and interpretability make them valuable tools for understanding the relationships between predictor variables and electricity consumption.

The MLP model, while demonstrating higher MSE compared to ensemble methods, still showcases commendable performance with an MSE of 99761.29. Its $R2$ score of 0.93 indicates its ability to explain approximately 93% of the variance in the data, highlighting its effectiveness in capturing underlying patterns in electricity consumption. MLPs, owing to their flexibility and capacity to discern complex nonlinear relationships, offer a balance between model complexity and interpretability.

Zone 4

The presented model performance metrics offer a comprehensive assessment of various machine learning algorithms applied to electricity consumption forecasting in Zone 4. These metrics, including Mean Squared Error (MSE) and R-squared ($R2$), serve as quantitative indicators of the predictive accuracy and explanatory power of each model.

Commencing with ensemble methods, XGBoost and Random Forest, both demonstrate competitive performance with relatively low MSE values of 20158.47 and 19936.25, respectively. These values indicate the algorithms' efficacy in minimizing the discrepancies between predicted and observed values. Moreover, achieving high $R2$ scores of 0.97 for both models underscores their capacity to explain approximately 97% of the variance in electricity consumption data. This reflects robust predictive capabilities and a strong ability to capture underlying patterns in the dataset.

Transitioning to linear regression-based approaches, Linear Regression and Ridge Regression, we observe higher MSE values compared to ensemble methods. Linear Regression yields an MSE of 450567.82, while Ridge Regression exhibits a slightly lower MSE of 450678.91. Despite these elevated MSE values, both models achieve respectable $R2$ scores of 0.70 and 0.69, respectively. This suggests that linear regression-based models can elucidate approximately 70% of the variance in the data. While they may not match the predictive accuracy of ensemble methods, their simplicity and interpretability make them valuable for understanding the relationships between predictor variables and electricity consumption.

The MLP model, while demonstrating higher MSE compared to ensemble methods, still showcases commendable performance with an MSE of 100086.29. Its $R2$ score of 0.93 indicates its ability to explain approximately 93% of the variance in the data, highlighting its effectiveness in capturing underlying patterns in electricity consumption. MLPs, owing to their flexibility and capacity to discern complex nonlinear relationships, offer a balance between model complexity and interpretability.

Zone 5

Beginning with ensemble methods, XGBoost and Random Forest both demonstrate impressive performance with relatively low MSE values of 19785.62 and 19696.84, respectively. These values indicate the algorithms' effectiveness in minimizing the discrepancies between predicted and observed values. Additionally, achieving high $R2$ scores of 0.98 for both models underscore their capacity to explain approximately 98% of the variance in electricity consumption data, reflecting robust predictive capabilities and a strong ability to capture underlying patterns in the dataset.

Transitioning to linear regression-based approaches, Linear Regression and Ridge Regression, we observe higher MSE values compared to ensemble methods. Linear Regression yields an MSE of 449950.13, while Ridge Regression exhibits a slightly lower MSE of 450089.25. Despite these elevated MSE values, both models achieve respectable $R2$ scores of 0.71 and 0.70, respectively. This suggests that linear regression-based models can elucidate approximately 71% of the variance in the data. While they may not match the predictive accuracy of ensemble methods, their simplicity and interpretability make them valuable for understanding the relationships between predictor variables and electricity consumption.

The MLP model, while demonstrating higher MSE compared to ensemble methods, still showcases commendable performance with an MSE of 99357.74. Its $R2$ score of 0.94 indicates its ability to explain approximately 94% of the variance in the data, highlighting its effectiveness in capturing underlying patterns in electricity consumption. MLPs, owing to their flexibility and capacity to discern complex nonlinear relationships, offer a balance between model complexity and interpretability.

Overall, the results emphasize the importance of selecting appropriate modelling techniques based on the specific characteristics of the dataset and the desired level of predictive accuracy. While ensemble methods like XGBoost and Random Forest excel in capturing complex relationships and achieving high accuracy, linear regression-based approaches offer simplicity and interpretability. MLPs provide a flexible and powerful modelling framework that can capture intricate patterns in the data. Understanding the trade-offs and strengths of each algorithm is crucial for making informed decisions in predictive modelling tasks, especially in domains like energy consumption forecasting where accurate predictions are essential for effective resource management and decision-making.

This paper embarks on a comprehensive journey into the realm of electricity consumption forecasting in Norway, a task of critical importance in the modern era of energy management and sustainability. Through meticulous analysis and experimentation, the study aims to unveil the intricate patterns and dynamics underlying electricity demand across the country's five bidding zones. Leveraging a diverse arsenal of predictive modeling techniques, including ARIMA, Random Forest, Linear Regression, Ridge Regression, MLP, and XGBoost, the research delves deep into the nuances of short-term and long-term load forecasting with a granularity of one-hour intervals.

The journey begins with an exploration of the ARIMA model, a stalwart of time-series analysis renowned for its simplicity and effectiveness in capturing temporal dependencies. Through rigorous evaluation and meticulous parameter tuning, the ARIMA model demonstrates exceptional predictive accuracy across all bidding zones, laying a solid foundation for further analysis.

However, the exploration does not end here; rather, it serves as a springboard into the realm of ensemble learning and neural network architectures. Ensemble methods like Random Forest and XGBoost emerge as formidable contenders, harnessing the power of decision trees and boosting techniques to capture complex relationships and achieve unprecedented levels of accuracy. Meanwhile, neural networks, represented by the Multilayer Perceptron (MLP), offer a flexible and powerful modeling framework capable of capturing intricate patterns in the data.

As the study unfolds, it becomes evident that each modeling technique brings its unique strengths and weaknesses to the table. Linear regression-based approaches offer simplicity and interpretability, making them valuable tools for baseline modeling and model interpretation. Ensemble methods excel in capturing complex relationships and achieving high accuracy, while neural networks provide a flexible framework for capturing intricate patterns in the data. Understanding the trade-offs and strengths of each algorithm is crucial for making informed decisions in predictive modeling tasks, especially in domains like energy consumption forecasting where accurate predictions are essential for effective resource management and decision-making.

This study offers a nuanced understanding of electricity consumption forecasting in Norway, showcasing the strengths and weaknesses of various predictive algorithms. While the ARIMA and XGBoost models emerge as a frontrunner in accuracy and reliability, the diverse performance of other models underscores the complexity of energy demand dynamics and the importance of tailored modeling approaches. As the energy landscape continues to evolve, insights from this study will serve as a guiding light for future research and decision-making in energy forecasting and resource management.

Chapter 6: Conclusion.

The journey embarked upon in this paper has been one of meticulous exploration and innovation, aimed at unlocking the secrets hidden within the intricate patterns of electricity consumption in Norway. Through the lens of various predictive modeling techniques, ranging from traditional time-series analysis to state-of-the-art machine learning algorithms, we have endeavored to unravel the complexities of energy demand forecasting with unprecedented accuracy and precision.

At the heart of our methodology lies the venerable ARIMA model, a stalwart of time-series analysis renowned for its simplicity and robustness. Through rigorous evaluation and meticulous parameter tuning, we have demonstrated the remarkable predictive prowess of ARIMA across diverse geographical zones in Norway. Its ability to capture both short-term fluctuations and long-term trends in electricity consumption with exceptional fidelity underscores its status as a cornerstone tool for energy forecasters.

Yet, our exploration did not end with ARIMA; rather, it marked the beginning of a journey into the realm of ensemble learning and neural network architectures. Leveraging the power of XGBoost, Random Forest, and Multilayer Perceptron (MLPs), we embarked on a quest to push the boundaries of predictive accuracy even further. The results were nothing short of extraordinary, with ensemble methods showcasing their prowess in capturing complex relationships and achieving unprecedented levels of accuracy, often surpassing the 99% mark.

However, amidst the excitement of technological innovation, we must not lose sight of the fundamental principles that underpin our predictive endeavors. Linear regression, with its elegant simplicity and interpretability, serves as a poignant reminder of the importance of understanding the trade-offs inherent in predictive modeling. While it may not boast the flashy performance metrics of its more sophisticated counterparts, its role in providing a baseline for comparison and aiding in model interpretation cannot be overstated.

As we reflect on the culmination of our efforts, it becomes evident that the true power of predictive modeling lies not in the superiority of any single algorithm, but in the synergy achieved through their judicious combination. Feature engineering, ensemble learning, and neural network architectures represent but a few pieces of the puzzle, each contributing to the tapestry of knowledge that drives innovation in energy consumption forecasting.

In conclusion, this study serves as a testament to the efficacy of diverse modeling techniques in forecasting electricity consumption across diverse geographical zones. From the venerable halls of ARIMA to the cutting-edge realms of ensemble learning and neural networks, our journey has been one of discovery, innovation, and relentless pursuit of excellence. As we peer into the future, the lessons learned from this endeavor will undoubtedly serve as guiding beacons for researchers and practitioners alike, shaping the trajectory of predictive modeling in energy forecasting for years to come.

Limitations and Future Research

While the paper presents a comprehensive approach to forecasting electricity consumption in Norway, it is important to recognize several limitations that may impact the generalizability and applicability of the findings. These limitations encompass various aspects, including data, methodology, and practical implications.

Firstly, the study's reliance on historical electricity consumption data from Norway raises concerns about its generalizability to other regions or countries with different socio-economic, geographical, and climatic conditions. Electricity consumption patterns are influenced by numerous factors such as population density, industrial activities, weather variations, and cultural behaviours, which may vary significantly across different regions. Therefore, extrapolating the findings of this study to other contexts should be approached with caution, and further validation using data from diverse locations is necessary to assess the model's robustness and transferability.

Secondly, while the paper employs a diverse set of modelling techniques, including ARIMA, regression models, ensemble methods, and neural networks, the choice of algorithms may not be exhaustive. There exists a wide array of advanced machine learning algorithms and forecasting methodologies that could potentially enhance predictive performance further. For instance, deep learning architectures like recurrent neural networks (RNNs) or long short-term memory (LSTM) networks have shown promising results in time-series forecasting tasks, particularly for capturing temporal dependencies and non-linear patterns in sequential data. Incorporating such advanced techniques could potentially yield even better predictions and uncover hidden insights in the electricity consumption data.

Another limitation pertains to the evaluation metrics used to assess the performance of the forecasting models. While the paper mentions achieving high accuracy rates, it is crucial to consider additional metrics such as mean absolute error (MAE), mean squared error (MSE), or root mean squared error (RMSE) to provide a comprehensive understanding of the model's predictive capabilities. Moreover, assessing the models' performance across different time horizons (e.g., short-term vs. long-term forecasts) and under varying conditions (e.g., peak vs. off-peak periods) would offer insights into their robustness and reliability in practical applications.

Furthermore, the paper discusses the importance of feature engineering, parameter tuning, and model optimization techniques in improving predictive accuracy. However, it lacks detailed insights into the specific feature selection criteria, hyperparameter optimization strategies, and validation procedures employed during the model development process. Providing a thorough explanation of these aspects would enhance the reproducibility and transparency of the study's findings, allowing other researchers to replicate and validate the results effectively.

Finally, while the study emphasizes the significance of accurate electricity consumption forecasting for resource management and decision-making, it does not delve into the practical implications and potential socio-economic impacts of the proposed models. Understanding how stakeholders, such as energy providers, policymakers, and consumers, can leverage these forecasting insights to optimize energy production, distribution, and consumption strategies is essential for realizing the full potential of the predictive models in real-world settings.

In conclusion, while the paper makes significant strides in forecasting electricity consumption in Norway using a diverse set of modelling techniques, it is essential to acknowledge and address the limitations to ensure the reliability, generalizability, and practical utility of the findings in broader contexts. Further research efforts focusing on data diversity, algorithmic advancements, comprehensive evaluation metrics, transparent methodology, and real-world applications are warranted to advance the field of energy consumption forecasting.

Uncategorized References

Badar, u. I., & Shams, F. A. (2022). Short-Term Electrical Load Demand Forecasting Based on LSTM and RNN Deep Neural Networks. *Hindawi*, *2022*(special). https://www.hindawi.com/journals/mpe/si/896958/

Banik, R., Das, P., Ray, S., & Biswas, A. (2021). Prediction of electrical energy consumption based on machine learning technique. *Electrical engineering*, *103*(2), 909-920. https://doi.org/10.1007/s00202-020-01126-z

Chicago Mercantile Exchange (CME) Group. *Understanding Basics of the Power Market*. https://www.cmegroup.com/education/courses/introduction-to-energy/introduction-to-power/understanding-basics-of-the-power-market.html

Demanuele C., Tweddell T., & Davies M. (2010). Bridging the gap between predicted and actual energy performance in schools. *In Proceedings of the World Renewable Energy Congress XI, Abu Dhabi, United Arab Emirates, 25–30 September 2010;*, 25-30.

Egging, R., & Tomasgard, A. (2018). Norway's role in the European energy transition. *Energy Strategy Reviews*, *20*, 99-101. https://doi.org/10.1016/j.esr.2018.02.004

Gang, C., Qingchang, H., Jin, W., Xu, W., & Yuyu, Z. (2023). Machine-Learning-Based Electric Power Forecasting. *Sustainability*, *15*. https://doi.org/10.3390/su151411299

Gievska, S., & Madjarov, G. (2019). Electrical Energy Consumption Prediction Using Machine Learning. *1110*, 72-82. https://doi.org/10.1007/978-3-030-33110-8_7 (Communications in Computer and Information Science)

Heba, A. S., Ashraf, W., & Adel, M. S. (2023). Smart Electricity Meter Load Prediction in Dubai Using MLR, ANN, RF, and ARIMA. *electronics*, *12*(2). https://www.mdpi.com/2079-9292/12/2/389

IEA, I. E. A. (2022). Norway Electricity Security Policy. https://www.iea.org/articles/norway-electricity-security-policy

Kang, J., & Reiner, D. M. (2021). What is the effect of weather on household

electricity consumption? Empirical evidence

from Ireland. *Energy Policy Research Group*, *2112*. https://www.econ.cam.ac.uk/cwpe

Khan, P. W., Kim, Y., Byun, Y.-C., & Lee, S.-J. (2021). Influencing Factors Evaluation of Machine Learning-Based Energy Consumption Prediction. *Energies (Basel)*, *14*(21), 7167. https://doi.org/10.3390/en14217167

Madeline, H. L. L., Yee, C. S., Ganeshsree, S., Pham, H. T., Le Cuong, L. H. S., Nguyen, T. T., & Vassilis, C. G. (2022). A Comparative Study of Forecasting Electricty ConsumptionUsing Machine Learning Models. *Mathematics*, *10*(1329). https://doi.org/10.3390/math10081329

Moral-Carcedo, J., & Vicens-Otero, J. (2005). Modelling the non-linear response of Spanish electricity demand to temperature variations. *Energy Economics*, *27*(3), 477-494. https://doi.org/10.1016/j.eneco.2005.01.003

Norby, T., Jensen, E. H., & Sartori, S. (2019). Deep decarbonization efforts in Norway for energy sustainability. *MRS Energy & Sustainability*, *6*, E10, Article E10. https://doi.org/10.1557/mre.2019.12

NordPool. *Bidding areas*. https://www.nordpoolgroup.com/en/the-power-market/Bidding-areas/

NVE, N. E. R. A. (2016). *Norway and the European power market*. https://www.nve.no/norwegian-energy-regulatory-authority/wholesale-market/norway-and-the-european-power-market/

Statista. (2020). *Electricity in Norway - statistics & facts*. https://www.statista.com/topics/9216/electricity-in-norway/#topicOverview

Statnett. Data from the power system In T. p. situation (Ed.).
Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, *32*(9), 1761-1768. https://doi.org/https://doi.org/10.1016/j.energy.2006.11.010
Vosooghzadeh, B. (2021). *Energy Consumption Theory*. https://doi.org/10.13140/RG.2.2.18889.06240
Wang, Z., Hong, T., Li, H., & Ann Piette, M. (2021). Predicting city-scale daily electricity consumption using data-driven models. *Advances in Applied Energy*, *2*, 100025. https://doi.org/https://doi.org/10.1016/j.adapen.2021.100025