## A black background with white text Description automatically generated with low confidence

MSc data science project

7PAM2002-0206-2023

Department of Physics, Astronomy and Mathematics

**Data Science FINAL PROJECT REPORT**

**Project Title:**

**Forecasting Electricity Demand in the United Kingdom.**

:

DANIEL Ndubisi Eze-22009281

Supervisor: Micheal Kuhn

Date Submitted: 5th May 2024

Word Count: ……..

# **DECLARATION STATEMENT**

# This report is submitted in partial fulfillment of the requirements for the degree of Master of Science in Data Science at the University of Hertfordshire.

# I have read the guidance to students on academic integrity, misconduct, and plagiarism information at [Assessment Offences and Academic Misconduct](https://www.herts.ac.uk/__data/assets/pdf_file/0007/237625/AS14-Apx3-Academic-Misconduct-v17.0.pdf) and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project module or course.

# I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged (Ref. UPR AS/C/6.1, section 7 and UPR AS/C/5, section 3.6). I have not used chatGPT or any other generative AI tool to write the report or code (other than where I have declared or referenced it).

# I did not use human participants or undertake a survey in my MSc project.

# 

# I hereby give permission for the report to be made available on module websites provided the source is acknowledged.

# 

# 

# Student name printed: **Daniel Ndubisi Eze**

# Abstract

# Contents

1. [Introduction 4](#_bookmark0)
2. [Background & Literature Review 6](#_bookmark1)

[3: Method 9](#_bookmark2)

* 1. [: Data Section 9](#_bookmark3)
     1. [: Data overview 9](#_bookmark4)

[3.1.1: Review of columns 9](#_bookmark5)

* + 1. [: Data type and quality 11](#_bookmark6)
    2. [: Data exploration and feature engineering 12](#_bookmark7)
  1. [Methodology: Choice of techniques 21](#_bookmark8)

[3 2.1: Method 1: Test encoding methods and use optimally encoded data with an advanced algorithm 21](#_bookmark9)

[3.2.2: Method 2: Use algorithm on raw data 24](#_bookmark10)

* 1. [: Metrics to measure success 25](#_bookmark11)

1. [Implementation & Results 27](#_bookmark12)
   1. [: Implementation of the different encoding techniques 27](#_bookmark13)
   2. [XGBoost implementation and results 28](#_bookmark14)
   3. [: Catboost implementation and results 31](#_bookmark15)
   4. [: Comparing the 2 algorithms 34](#_bookmark16)
2. [Analysis 35](#_bookmark17)
   1. [: Putting the results into perspective 35](#_bookmark18)
   2. [: Which is the best model given the objectives of the project? 36](#_bookmark19)
   3. [: Further analysis of the XGBoost Classifier’s results 36](#_bookmark20)
   4. [: Analysing the total approach 37](#_bookmark21)
   5. [: The CatBoost Classifier had poor results 38](#_bookmark22)
   6. [: Feedback from OLBG 38](#_bookmark23)
3. [Conclusion 39](#_bookmark24)
   1. [: Summary of findings 39](#_bookmark25)
   2. [: Ideas for future work 39](#_bookmark26)
4. [References 41](#_bookmark27)
5. [Appendix 43](#_bookmark28)

# Introduction

At the forefront of the United Kingdom’s ambitious journey towards a sustainable and secure energy landscape, the art and science of forecasting electricity demand emerges as pivotal. This complex challenge, integral to achieving the nation’s environmental goals, economic resilience, and enhancing the daily lives of its citizens, has never been more significant. As the UK boldly navigates its transition to renewable energy and commits to carbon neutrality, the accuracy of electricity demand forecasting becomes a linchpin in this transformative era.

The task at hand is far from straightforward. The integration of renewable energy sources like wind and solar introduces a layer of unpredictability, while technological leaps, from electric vehicles (EVs) to smart home innovations, are reshaping consumption patterns. These evolutions necessitate a forecasting approach that is both sophisticated and adaptable, capable of ensuring the reliability of the energy system amidst rapid change.

This narrative delves into the quest for innovative forecasting methodologies that cater to the critical needs of the National Grid. Accurate forecasting stands as a cornerstone for balancing supply and demand, optimizing costs, and ensuring the safety and reliability of the energy supply. By pushing the boundaries of traditional models with advanced technologies and insights, this project aims not just to refine forecasting accuracy but to redefine the blueprint of the UK’s energy future.

As I embark on this exploration, I aim to uncover the intricate dance of predicting electricity demand—a journey from the foundational role of the National Grid Electricity System Operator (ESO) in real-time balancing to the impacts of digital advancements. Additionally, to create a forecasting model that not only aligns with current requirements but also illuminates the path towards environmental objectives, reliable energy access, and resilient infrastructure.

This project will initiate a discussion on the challenges, methodologies, and importance of accurate electricity demand forecasting.There will be recognition that this endeavor is more than just a technical pursuit. The goal is to craft a sustainable, efficient, and secure energy roadmap for future generations.

# Importance of Promoting Energy Efficiency and Sustainability

The heart of sustainable energy management practices is the imperative to use energy resources judiciously and efficiently. Effective demand forecasting is instrumental in achieving these goals, as it provides a roadmap for incorporating renewable energy sources into the grid. By anticipating demand fluctuations, energy providers can optimize the mix of renewable and non-renewable energy sources, thus reducing reliance on fossil fuels and minimizing the environmental footprint of energy production.

Furthermore, the ability to predict electricity demand with high accuracy is a key factor in the development of energy efficiency measures. It informs the creation and implementation of policies aimed at reducing energy consumption without compromising the quality of service. This is particularly relevant in the context of global efforts to combat climate change, where energy efficiency is recognized as a critical element in the transition towards a low-carbon economy.

## Role in Operation and Planning of the Electricity Grid

The operation of the electricity grid is a complex and dynamic task, requiring a delicate balance between the generation of electricity and its consumption by end-users. Forecasting plays a vital role in this process, enabling grid operators to manage the flow of electricity efficiently and to ensure that the grid remains stable and reliable. Accurate demand forecasts are essential for planning maintenance activities, managing peak load periods, and avoiding blackouts or other disruptions in service.

Additionally, the planning aspect of grid management benefits significantly from precise demand forecasting. It allows for the strategic expansion of grid infrastructure, the integration of smart grid technologies, and the development of demand response strategies. These efforts collectively contribute to the grid's resilience and its ability to meet current and future energy needs.

## Enabling the Integration of Renewable Energy Sources

The integration of renewable energy sources into the electricity grid is a critical component of the transition to a more sustainable and environmentally friendly energy system. Demand forecasting is crucial in this context, as it helps to address the inherent variability of renewable energy sources such as wind and solar power. By accurately predicting demand, energy providers can better match the supply of renewable energy with consumption patterns, thereby maximizing the utilization of these clean energy sources.

In summary, forecasting electricity demand is an indispensable activity in the energy sector, with profound implications for energy efficiency, sustainability, and the operational reliability of the electricity grid. It embodies a proactive approach to energy management, enabling the sector to navigate the challenges of a rapidly changing energy landscape. As we delve deeper into the factors influencing electricity demand in the UK, it becomes evident that this task is influenced by a myriad of factors, each adding layers of complexity to the forecasting process.

# Factors Influencing Electricity Demand in the UK

The demand for electricity in the United Kingdom is shaped by a complex interplay of factors, each contributing to the fluctuating nature of consumption patterns. Understanding these influences is essential for accurate demand forecasting, enabling energy providers to adapt their strategies and ensure a stable and efficient supply of electricity. The primary factors influencing electricity demand in the UK include seasonal and weather patterns, economic growth and industrial activity, population growth and urbanization, and technological advances and energy efficiency measures.

## Seasonal and Weather Patterns

One of the most significant determinants of electricity demand in the UK is the variation in seasonal and weather conditions. During the winter months, the demand for electricity surges as colder temperatures necessitate increased heating requirements. This spike in demand is further accentuated on particularly cold days, where additional electricity is consumed to maintain comfortable indoor temperatures. Conversely, during milder seasons, the demand for electricity decreases as the need for heating diminishes. This seasonal variability requires energy providers to anticipate and plan for peak demand periods, ensuring that sufficient capacity is available to meet the increased load.

## Economic Growth and Industrial Activity

The level of economic activity within the UK directly influences electricity consumption patterns. Periods of economic growth are typically associated with increased industrial production, commercial activities, and overall energy consumption. As businesses expand and industrial operations intensify, the demand for electricity rises to power machinery, equipment, and other energy-intensive processes. Conversely, during economic downturns, reduced industrial activity leads to a decrease in electricity demand. Accurately forecasting these economic fluctuations is crucial for managing electricity supply effectively.

## Population Growth and Urbanization

Population growth and the expansion of urban areas have a direct impact on electricity demand in the UK. As the population increases, so does the need for residential housing, commercial spaces, and public services, all of which require electricity. Urbanization, in particular, drives the concentration of energy consumption within cities, where the density of population and economic activities heightens the demand for electricity. This trend underscores the importance of incorporating demographic and urban development data into demand forecasting models.

## Technological Advances and Energy Efficiency Measures

Advancements in technology and the implementation of energy efficiency measures can influence electricity demand in two ways. On one hand, the introduction of energy-efficient appliances, lighting, and industrial equipment can lead to a reduction in overall electricity consumption. These technologies improve the efficiency of energy use, thereby decreasing the demand for electricity. On the other hand, the proliferation of new technologies, such as electric vehicles and digital devices, can increase electricity consumption. The adoption rates of these technologies and their impact on demand vary, presenting both opportunities and challenges for demand forecasting.

**Aims**

This project aims to develop a forecasting model that can accurately estimate future electricity requirements. This will provide valuable insights for informed energy policy, infrastructure optimization, and achieving environmental goals, while ensuring reliable energy access throughout the UK.

**Objectives**

To identify advancements in forecasting methodologies that can address the unpredictability of renewable energy sources and the evolving pattern of electricity demand.

To recommend areas for future research and investment to support the development of more sophisticated demand forecasting tools.

# Background & Literature Review

Forecasting electricity demand is a cornerstone activity in the management of the energy sector, reflecting its critical role in underpinning energy efficiency, sustainability, and security. This process goes beyond mere prediction, serving as a fundamental mechanism for the effective operation and strategic planning of the electricity grid. It facilitates the seamless integration of renewable energy sources, ensuring that energy consumption needs are met with precision and reliability. At its core, the forecasting of electricity demand is pivotal for maintaining an equilibrium between supply and demand, thus preventing wastage and guaranteeing that energy is accessible when and where it's needed most.

The essence of forecasting electricity demand lies in its ability to inform and guide the energy sector towards making informed decisions regarding the generation, distribution, and consumption of electricity. This foresight is invaluable, as it enables energy providers to prepare for and adapt to varying consumption patterns, thereby enhancing the efficiency of the grid's operation. Moreover, accurate demand forecasting plays a crucial role in the economic aspect of energy management, influencing investment decisions, operational costs, and ultimately, the pricing of electricity for consumers

# Challenges in Forecasting Electricity Demand

Accurately estimating future electricity demand in the UK is fraught with challenges. High-quality, detailed data is essential for precise forecasting, yet such data can be difficult to obtain. Rapid technological advancements and shifts in consumer behavior add layers of complexity to the forecasting process, necessitating sophisticated models that can adapt to these changes. Furthermore, the integration of renewable energy sources introduces variability that complicates demand predictions. Regulatory and policy changes can also have unforeseen impacts on electricity demand, requiring forecasting models to be flexible and adaptable.

## Impact of Accurate Electricity Demand Forecasts

The benefits of precise electricity demand forecasting are manifold, affecting areas such as energy policy and planning, economic stability, environmental sustainability, and grid efficiency. Accurate forecasts enable energy providers to optimize infrastructure investments, reduce operational costs, and facilitate the transition to a low-carbon economy. Moreover, reliable demand predictions are crucial for maintaining grid stability, enhancing the efficiency of electricity distribution, and ensuring a balance between supply and demand.

Forecasting electricity demand in the UK is a multifaceted challenge with far-reaching implications for energy policy, economic stability, and environmental sustainability. The next section will delve into the specific methodologies and analyses presented in various studies, highlighting the advancements and limitations in current forecasting models.

# Review of Forecasting Methods and Papers

The task of forecasting electricity demand in the UK and beyond has prompted a diverse array of approaches, each reflecting different assumptions, strengths, and limitations. This section reviews several key studies that have contributed to the field, offering insights into the evolution of forecasting methodologies.

**Bianco, Manca, and Nardini (2009)**

This study utilizes a linear regression approach to model electricity demand, emphasizing simplicity and accessibility in forecasting. The authors argue for the model's foundational value, highlighting its utility in capturing broad trends in electricity usage. However, the study acknowledges a significant limitation: the assumption of a static linear relationship between variables. This assumption overlooks the dynamic and fluctuating nature of electricity demand, particularly influenced by external factors such as weather and seasonal changes. The study points to the necessity for more complex, adaptable models capable of accounting for these variations.

**De Falco, Di Noia, and Rizzo (2021)**

Focusing on the photovoltaic power sector, De Falco and colleagues explore the exponential smoothing method, which assigns exponentially decreasing weights to past observations. This method, especially in its triple exponential form, is lauded for its ability to incorporate trend and seasonality into forecasts. However, the study also identifies a crucial drawback: the method's assumption of independent random noise. In reality, the noise within electricity demand data often carries intrinsic correlations, challenging the exponential smoothing method's effectiveness in capturing seasonal and trend-related patterns accurately.

**Chaturvedi et al. (2022)**

By comparing SARIMA, LSTM, and Prophet models, Chaturvedi et al. delve into the frontier of forecasting techniques. Their analysis demonstrates these models' capability to grasp complex nonlinear relationships and seasonal patterns within electricity demand data. The study underscores the computational intensity and sophistication required to leverage these models effectively, pointing to a significant evolution in the analytical approaches to demand forecasting. This comparison highlights the importance of selecting appropriate models that align with the specific characteristics and complexities of the data in question.

**Barassi, M., & Zhao, Y. (2018)**

Barassi and Zhao adopt a comprehensive approach by integrating various forecasting techniques to enhance the accuracy of short-term energy demand forecasts. Their methodology involves a combination of linear and non-linear models, including ARMA, Holt-Winters, and various neural network approaches. The innovative aspect of their study lies in the use of an algorithm that selects the optimal forecasting model based on specific criteria. This approach, coupled with the application of multiple weighting metrics for model averaging, represents a sophisticated strategy aimed at improving forecast reliability and adaptability.

**Hu et al. (2019)**

Hu and colleagues highlight the effectiveness of machine learning models in forecasting electricity demand while also acknowledging their limitations in providing insights into the causes of structural changes. They advocate for the use of hybrid models, which combine different forecasting approaches to enhance accuracy and interpretability. This perspective is particularly valuable in addressing the multifaceted nature of electricity demand, where understanding the underlying factors influencing changes is as important as predicting the demand itself.

# Critical Analysis and Relevance

In the realm of forecasting electricity demand, the critical analysis of various methodologies reveals a nuanced landscape where the choice of technique significantly impacts the accuracy and reliability of predictions. The studies reviewed encapsulate a spectrum of approaches, from linear regression to complex machine learning models, each with distinct advantages and challenges. This section delves into the implications of these methodologies for future research and the development of forecasting models, emphasizing the necessity for continuous innovation and adaptation in response to the evolving energy landscape.

Adaptation to Dynamic Energy Consumption Patterns

The dynamic nature of electricity demand, influenced by factors such as technological advancements, policy changes, and shifts in consumer behavior, demands forecasting models that are not only accurate but also adaptable. The transition from simpler models, like linear regression, to more sophisticated ones, such as LSTM and hybrid models, reflects an ongoing effort to better capture the complexities of energy consumption. Future research should focus on enhancing the flexibility of these models to accommodate rapid changes in the energy sector, ensuring forecasts remain relevant and reliable.

Incorporating External Variables and Renewable Energy Integration

A critical aspect of forecasting electricity demand is the ability to account for external variables, such as weather conditions, economic activity, and the integration of renewable energy sources. The variability introduced by renewable energy, in particular, presents a significant challenge, necessitating models that can effectively handle fluctuations in supply and demand. Research efforts should continue to explore and refine methodologies that integrate these variables into forecasting models, improving their ability to predict demand under varying conditions.

Policy and Planning Implications

Accurate demand forecasts are essential for effective energy policy and infrastructure planning. They provide a foundation for decisions related to grid expansion, renewable energy integration, and investment in energy efficiency measures. As forecasting methodologies evolve, it is crucial that policy-makers and planners stay abreast of these advancements, leveraging the latest models to inform strategic decisions. Future research should also consider the policy implications of forecasting inaccuracies, exploring ways to mitigate risks and enhance the resilience of energy systems.

Bridging the Gap Between Accuracy and Interpretability

While advanced machine learning models offer improved accuracy in demand forecasting, they often lack interpretability, making it challenging to understand the underlying factors driving predictions. This gap highlights the need for research focused on developing models that balance accuracy with interpretability. Hybrid models, which combine machine learning with more traditional statistical approaches, offer a promising avenue for achieving this balance. Future studies should explore these hybrid approaches further, enhancing the ability of forecasters to provide insights into the causes of demand fluctuations.

# 

# Dataset:

# Source

# The dataset was obtained from Kaggle and compiled by the National Grid Electricity System Operator (ESO) for Great Britain, which is responsible for electricity system operations in the country.The data is publicly available and has been regularly updated since its initial collection in 2009. This extensive dataset provides a detailed record of electricity demand across Great Britain, updated every half hour, translating into 48 data points per day. It comprises 4,978,950 entries organized into 262,050 rows and 19 columns.

# Collection Methodology and Purpose

# National Grid ESO employs an advanced metering infrastructure to accurately record electricity demand across Great Britain. This system captures data at thirty-minute intervals, offering a high-resolution view of energy consumption patterns over time. The primary purpose of collecting this data is to facilitate the efficient operation of the electricity system, ensuring stability and reliability. Additionally, the data serves a crucial role in energy forecasting, policy making, and academic research, providing a foundational resource for understanding and analyzing electricity demand dynamics.

# Dataset Composition

# The dataset is composed of several files, each serving different analytical purposes:

# Annual Demand Files (Historic\_demand\_year\_20xx.csv): Contain detailed demand data for each year.

# Merged Annual Dataset (Historic\_demand\_year\_2009\_2024.csv): Combines all yearly data into a single file for longitudinal analysis.

# Cleaned Merged Dataset (Historic\_demand\_year\_2009\_2024\_noNaN.csv): Offers a cleaned version of the merged dataset with NaN values removed and includes hourly data points for more granular analysis.

# **The dataset includes columns**:

|  |  |
| --- | --- |
| **settlement\_date** | **The date on which the data was recorded, formatted as dd/mm/yyyy.** |
| **settlement\_period** | **The half-hourly interval of the data record, indicating the specific time period within the day.** |
| **period\_hour** | **hour of the day for the settlement period, enhancing the temporal granularity of the data.** |
| **nd** | **The total electricity demand across Great Britain, excluding certain uses, measured in megawatts (MW).** |
| **tsd** | **The ND plus the additional generation required to meet station load, pump storage pumping and interconnector exports. Measured in MW.** |
| **England\_wales\_demand** | **Total demand for England and Wales** |
| **embedded\_wind\_generation** | **An estimate of the electricity generated by wind farms not directly metered by the system, serving to reduce the apparent demand, measured in MW.** |
| **embedded\_wind\_capacity** | **The estimated total capacity of embedded wind generation facilities across Great Britain, measured in MW.** |
| **embedded\_solar\_generation** | **An estimate of the electricity generated by solar panels not directly metered by the system, also reducing the apparent demand, measured in MW.** |
| **embedded\_solar\_capacity** | **The estimated total capacity of embedded solar generation facilities across Great Britain, measured in MW.** |
| **non\_bm\_stor** | **Generation or demand reduction resources not included within the National Demand definition, measured in MW.** |
| **pump\_storage\_pumping** | **The electricity demand due to operation of hydroelectric pump storage units, where negative values indicate pumping, measured in MW.** |
| **ifa\_flow** | **The net electricity flow through the IFA interconnector, with negative values indicating exports from Great Britain, measured in MW.** |
| **ifa2\_flow** | **Similar to IFA\_FLOW, but for the IFA2 interconnector.** |
| **britned\_flow** | **The net electricity flow through the BritNed interconnector, with negative values indicating exports from Great Britain, measured in MW.** |
| **moyle\_flow** | **The net electricity flow through the Moyle interconnector, with negative values indicating exports from Great Britain, measured in MW** |
| **east\_west\_flow** | **The net electricity flow through the East-West interconnector, with negative values indicating exports from Great Britain, measured in MW.** |
| **nemo\_flow** | **The net electricity flow through the Nemo interconnector, with negative values indicating exports from Great Britain, measured in MW.** |
| **is\_holiday** | **An indicator of whether the data record corresponds to a public holiday, which can significantly impact electricity demand patterns (0 or 1)** |

# The key columns for this project in the dataset are: Settlement\_date and tsd(transmission\_system\_demand)

# Justification for Dataset Selection

# This dataset was chosen for its unparalleled depth and granularity, providing a unique opportunity to analyze electricity demand patterns with high temporal resolution. Such detailed data are essential for forecasting electricity demand.

# **Ethical Considerations**

# Given that the dataset aggregates national-level data without identifying individual consumption patterns, privacy concerns are minimal. The following displays the granted rights and conditions of use for the dataset. Based on this information, the dataset does not require ethical approval for its use.

# 

# 

# 

# **Exploratory data analysis**

# The dataset is loaded in pandas dataframe. Settlement\_date column was converted to datetime format and then indexed.Feature variable (‘transmission system demand; tsd’) was extracted and aggregated to daily values and interpolated. It is important to aggregated to daily values to minimise the effects of outliers and anamolies in the dataset while interpolating handles missing values. As shown below, the aggregated dataset.

**settlement\_date**

**2009-01-01 38528.395833**

**2009-01-02 41133.458333**

**2009-01-03 40667.791667**

**2009-01-04 41013.958333**

**2009-01-05 47322.312500**

**Freq: D, Name: tsd, dtype: float64**

# Fig.1 shows the distribution of aggregated data over time. The data seems to show a single clear peak which distribution occurring in a bin around the 30000 range. This indicates the most common range of transmission system in the dataset.

# The shape of the distribution is somewhat symmetric, with a slight skew to the right. This suggests that there are more instances of higher system demand values than lower ones, indicating a trend in the dataset.

# 

# **Figure 1: Distribution of the transmission system demand.**

# Fig. 2 shows the daily aggregated data line plot.This provides an overview of the trends and cyclical changes in the data over time. It can be inferred from the plot that there exists a clear seasonal pattern in the data, with peaks and troughs repeating approximately annually. The peaks indicate a high electricity demand during specific times of the year. This could be associated with heating in the United Kingdom during the winter period.

# 

# **Figure 2: Trend of transmission system demand over time.**

# To gain a better understanding of the data dynamics, underlying patterns, and modeling, I decomposed the aggregated data using the seasonal\_decompose function from the statsmodels library with an additive model. Fig. 3 displays the components of the decomposed data as trends, seasonalities, and residuals.

# This insight is important in forecasting by allowing models to account for and predict each component separately. It is also relevant to identify anomalies in the data, such as unusually high or low values, that do not fit the typical pattern.

# 

# **Figure 3: Decomposed aggregated data**

# Further analysis was conducted to extract yearly, monthly, weekly, and daily aggregated data over time.Fig. 4 shows the transmission system demand over different time aggregations. These graphs provide a multi-tiered view of the system demand, from high-level yearly trends to fine-grained daily variations. This could be useful for understanding overall trends, seasonal effects, weekly cycles, and daily fluctuations, which can inform demand planning, resource allocation, and system optimization strategies.

# 

# Figure 4: Transmission system demand over different time aggregations.

# 3: Methodology:

In my approach to forecasting system demand, I dedicate my efforts to meticulous data preparation, model selection, and evaluation with the aim of capturing the intricate patterns in transmission system demand and making accurate predictions.

The first step in my methodology involved importing the dataset into a Python environment using Jupiter Notebook. And preparing the time-series data for transmission system demand analysis. Due to computational limitations regarding the type/size of my dataset, it is important to aggregate it into daily values. To address missing values or anomalies in the data, linear interpolation was used, and any irregularities were smoothed out using the moving average technique. The data was then scaled to a standard range to facilitate more efficient learning by the neural network models.

Model Implementation

I chose three forecasting models for their distinct advantages and capabilities in time series analysis. The implementation was executed in Python, leveraging various libraries designed for statistical and machine learning tasks. The selected models are Prophet, LSTM and ARIMA.

**Prophet**

Prophet, developed by Facebook, is a decomposable time series model that handles trends, seasonality, and holidays. It is robust to missing data and shifts in trends, and typically works well with daily data that exhibit strong seasonal patterns. I used the Prophet model to accommodate nonlinear trends due to its flexibility in modeling nonstationary data with additive components. I also adjusted the model’s parameters and tailored it to capture the holiday effect in my dataset.

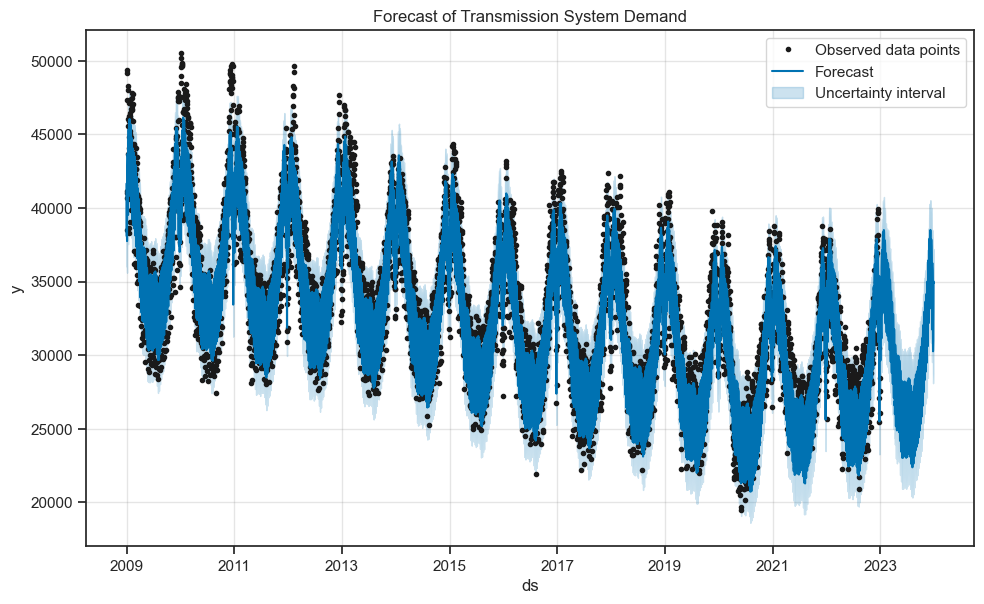
In general, the Prophet model can be represented as follows: **y(t)=g(t)+s(t)+h(t)+E(t)**

Where g(t) represents the trend function, s(t) represents periodic changes, h(t) represents the holiday effect, and E(t) represents the error term. It can be seen that the Prophet model is suitable for datasets that have trends, seasonalities, and holidays.

The aggregated daily transmission system demand data were split for training and testing. The model was not initialized with the prepared holidays dataframe and fit. I created a future dataframe of 365 days with daily frequency and make predictions.

Fig 5: shows the model plot of the forecast, which includes observed data points, the forecasted trend, and an uncertainty interval. The black dots represent the actual observed values of the TSD over time, from 2009 to beyond 2023. These points are critical as they serve as the ground truth for the model's forecast. The cyclical nature of these points suggests a strong seasonality within the data, as well as an overall variance that the forecasting model aims to capture. The solid blue line indicates the forecasted TSD values. This line provides a visual guide to the central tendency expected by the forecasting model, likely generated by a time series forecasting method such as Prophet, given the context of the question. The model seems to have effectively learned the seasonal peaks and troughs of the system demand, which it then projects forward into the future. The shaded blue area represents the uncertainty interval, often calculated as a credible or prediction interval. This interval offers a range within which future observations are expected to fall with a certain level of confidence. It appears to widen as we move further from the last observed data point, suggesting increasing uncertainty in the forecast the further out we project. This is common in time series forecasting, as predictions are generally less certain the farther they extend into the future.

The model appears to fit the historical data well, capturing the seasonal fluctuations. This is evidenced by the forecasted values aligning closely with the observed data points, and the uncertainty intervals encompassing the variability in the data. This implies that the model's parameters are well-tuned to the historical patterns.



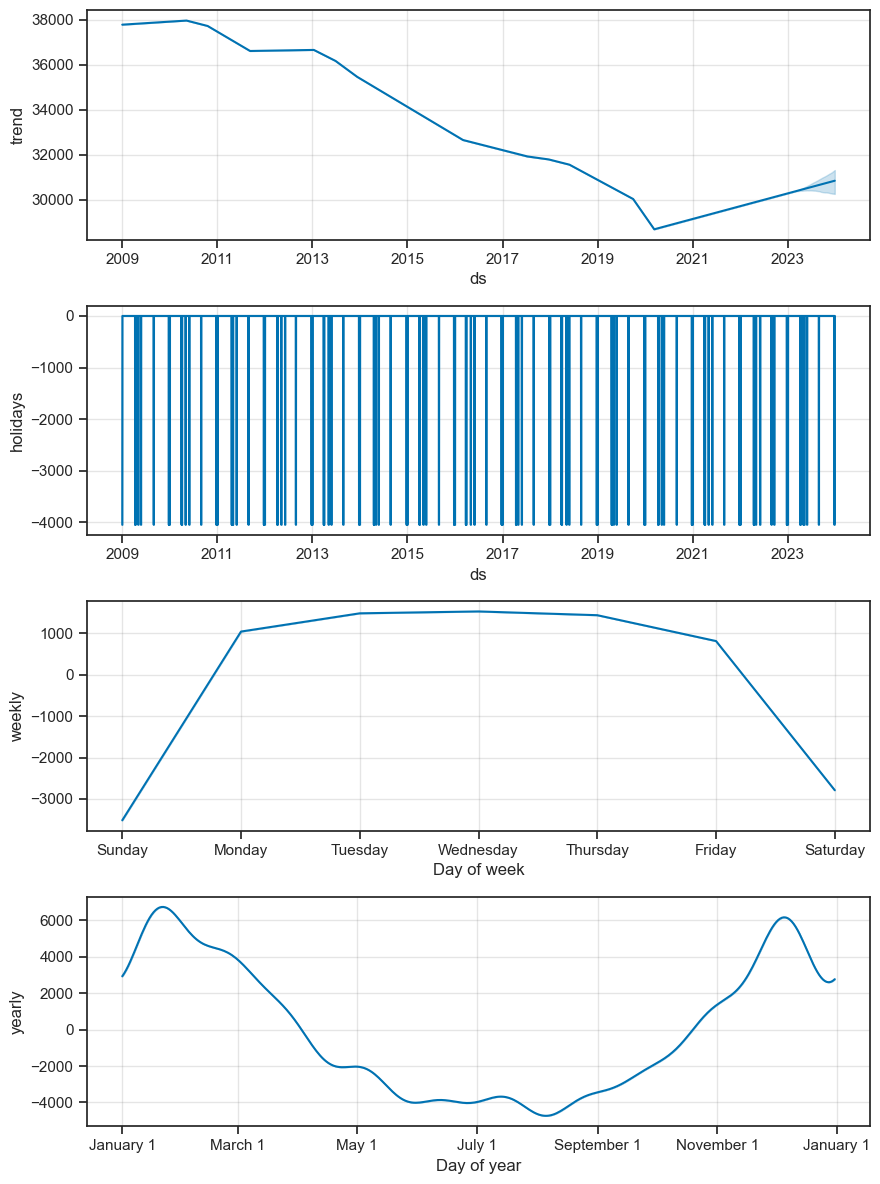
**Figure 5: Plot of the Prophet model forecast.**

I went further to plot the components of the model’s forecast. Remember that it is an addictive model. The essence is to see the components of the model’s forecast. These components are the form of trends, seasonalities, and holidays.

Fig. 6 displays these components, and it follows the same pattern as the extracted ‘tsd’ over different time aggregation In fig. 4. The components are trend, holidays, weekly, and yearly components.

The second graph displays the effects of holidays on system demand during the same period. The negative values indicate that on holidays, the system demand is consistently lower than on a typical day, which could be due to businesses being closed and industrial activities. The consistency of the negative effect across all years suggests that it remains relatively stable over time. The graph depicting weekly seasonalities illustrates the effect of weekly seasonality. It shows that system demand is highest at the start of the week, gradually decreasing as the week progresses, with the lowest demand on Saturdays. This weekly pattern may reflect the operational cycles of businesses and industries that follow a traditional workweek schedule. The bottom graph illustrates the yearly seasonality effect. There is a clear pattern that peaks around the beginning and the middle of the year. These peaks might correspond to extreme temperature seasons (like winter and summer) when heating and cooling demand drive higher energy consumption. There's also a notable dip in demand that occurs roughly between these two peaks, possibly corresponding to milder weather conditions when there is less need for heating or cooling.

When using these components for forecasting, it is important to consider that while some patterns may remain constant, others can change due to evolving social habits, technological advancements, economic factors, and policy changes. The slight uptick in the trend after 2021, for example, justifies need for further investigation.



**Figure. 6 components of the given forecast.**

The Prophet model makes forecasts with uncertainty intervals which represent the range of values within which the actual future values are expected to lie, with a certain probability. It is a statistically derived range that reflects the confidence in the model's forecasts based on the patterns found in the historical data. It's a vital tool for interpreting the forecasts and making informed decisions based on them. Understanding the bounds of the uncertainty interval is critical for risk management.

It helps in planning for best-case and worst-case scenarios. As shown below, the yhat\_lower and yhat\_upper are the uncertainty intervals from the Prophet model with defualt confidence interval of 80%.

The resulting plot is displayed in Fig. 7 below.

**yhat\_upper**

|  | **ds** | **yhat** | **yhat\_lower** |
| --- | --- | --- | --- |
| **5110** | **2022-12-29** | **34335.790581** | **32026.495970** | **36416.386953** |
| **5111** | **2022-12-30** | **33760.443978** | **31780.296796** | **35866.503662** |
| **5112** | **2022-12-31** | **30241.284779** | **28239.595864** | **32447.231050** |
| **5113** | **2023-01-01** | **29631.412128** | **27304.389144** | **31748.316248** |
| **5114** | **2023-01-02** | **30292.908565** | **28218.741038** | **32422.907657** |
| **...** | **...** | **...** | **...** | **...** |
| **5470** | **2023-12-24** | **30261.877463** | **28092.648062** | **32426.641332** |
| **5471** | **2023-12-25** | **30646.344914** | **28477.572490** | **32812.687098** |
| **5472** | **2023-12-26** | **30988.478548** | **28752.333124** | **33077.842764** |
| **5473** | **2023-12-27** | **35013.570609** | **32916.237014** | **37207.835425** |
| **5474** | **2023-12-28** | **34892.303991** | **32756.385569** | **37174.460952** |

**365 rows × 4 columns**

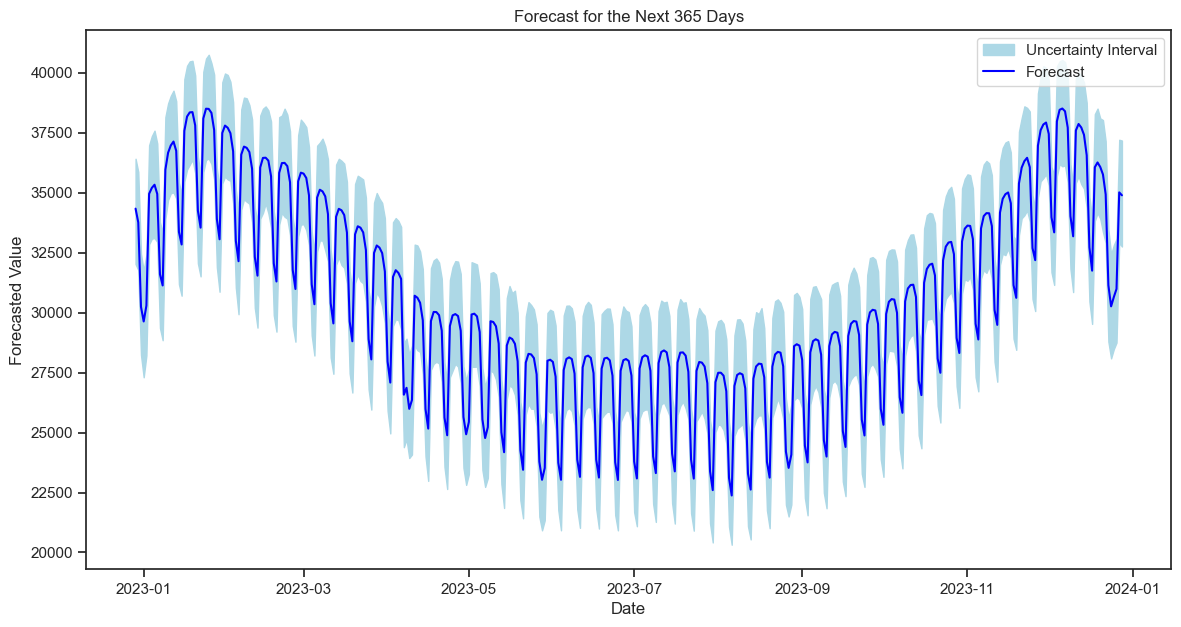


Figure 7. Forecast for the next year with uncertainty interval.

**Long Short-Term Memory (LSTM) networks.**

This is a type of recurrent neural network (RNN) that can learn long-term dependencies.They were introduced by Hochreiter & Schmidhuber (1997), and were developed to deal with the vanishing gradient problem that can occur when training traditional RNNs. The vanishing gradient problem involves the gradients shrinking and becoming too small for effective learning during the backpropagation process in deep neural networks.

For the LSTM model, I constructed a neural network architecture using TensorFlow and Keras, tuning the number of layers and neurons to optimize the learning from sequential data. Special attention was given to avoiding overfitting through the use of regularization techniques and dropout layers. Fig. 8 shows the architecture of the LSTM.

**Model: "sequential\_1"**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Layer (type) Output Shape Param #**

**=================================================================**

**lstm\_2 (LSTM) (None, 24, 50) 10400**

**dropout\_2 (Dropout) (None, 24, 50) 0**

**lstm\_3 (LSTM) (None, 50) 20200**

**dropout\_3 (Dropout) (None, 50) 0**

**dense\_1 (Dense) (None, 1) 51**

**=================================================================**

**Total params: 30651 (119.73 KB)**

**Trainable params: 30651 (119.73 KB)**

**Non-trainable params: 0 (0.00 Byte)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Figure 8. LSTM model architecture**

The sklearn.preprocessing library includes the MinMaxScaler function, which is used to preprocess data for input into the LSTM model. The aggregated daily data has been split into training, testing, and validation data. And applying normalization.

The number of data points for each split is determined by applying the defined proportions to the total dataset size. The training size is 70% of the total size.Validation size is 20% of the total size. The test size refers to the remaining portion of the dataset after accounting for the sizes of the training and validation sets.

A MinMaxScaler is used to normalize the dataset to a range of 0 to 1.

The scaler is fitted only on the training data to prevent information leakage from the validation and test sets.

After fitting, the scaler is then used to transform the validation and test sets, ensuring that all splits are on the same scale.

I built the model and trained it using 100 epochs. And plotted the loss metrics, as shown in Fig 9. The graph suggests a well-tuned training process, with the model showing signs of good generalization capabilities.

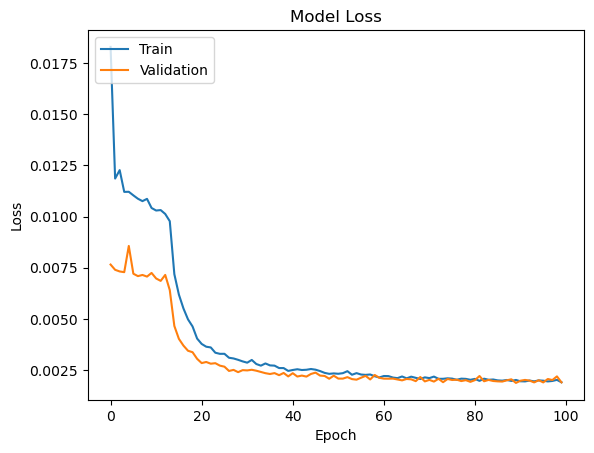
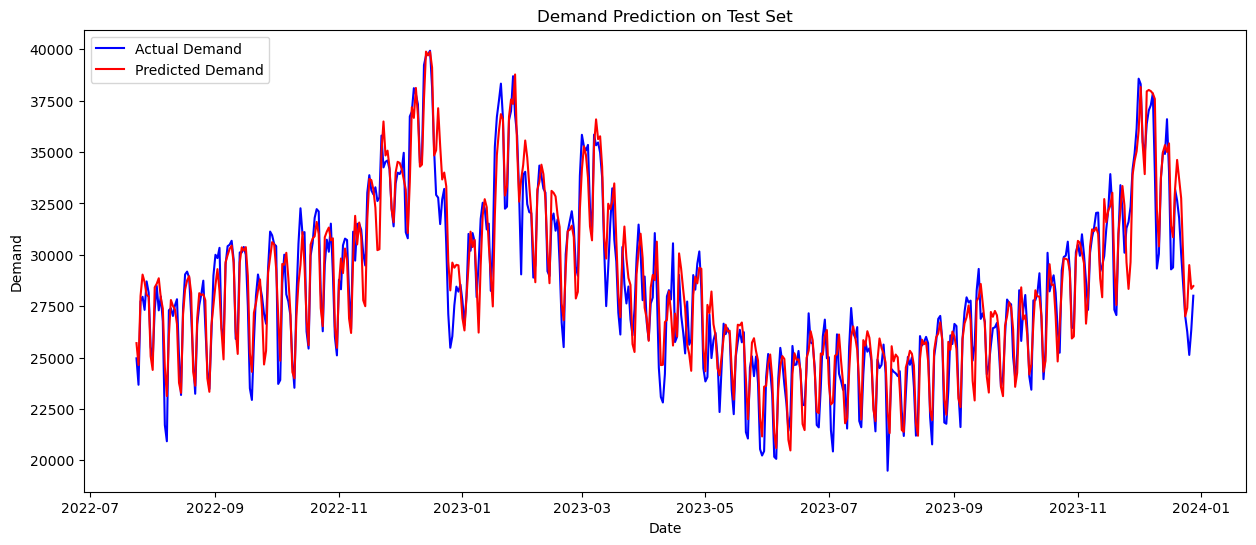


Figure. 9: LSTM loss metrics plot over 100 epochs.

Visual representation of the prediction on the test data is shown in Fig.10. The graph illustrates a close match between actual and predicted demand, highlighting the forecasting model over a period from July 2022 to January 2024. It shows a consistent declining trend in demand with clear seasonal patterns that the model successfully captures. The predictions closely mirror actual values, indicating the high precision and reliability of the model in forecasting electricity demand.



**Figure. 10 predictions on the test data.**

ARIMA

ARIMA was selected for its proficiency in modeling linear relationships. I determined the order of the model (p, d, q) through iterative testing and examination of the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots.

Model Training and Validation

Each model was trained on a dataset spanning several years, withholding the final year as a test set to evaluate forecasting performance.I applied time series cross-validation, progressively expanding the training dataset and testing on subsequent periods to ensure robustness.

Results

When evaluating the models, I chose multiple metrics to capture different aspects of forecasting performance:

MAE provided a straightforward measure of average absolute errors, giving me a practical sense of model accuracy.

RMSE was crucial for highlighting larger errors, as it assigns a higher penalty to large prediction errors, which is critical for system demand planning.

MAPE offered insights into the relative error, making it easier to communicate model performance in percentage terms.

To present my results, I compiled a table comparing the performance metrics of each model, alongside a set of graphs illustrating actual vs. predicted demand to visually assess the models' performance.

Analysis and Discussion

The results indicated that the LSTM network exhibited superior performance, likely due to its ability to capture complex patterns and dependencies over time. However, Prophet performed remarkably well during peak demand periods, attributed to its explicit modeling of yearly seasonality and holidays.

In comparing to existing literature, the performance of LSTM in my analysis aligns with recent studies highlighting the efficacy of neural networks in time series forecasting.

Limitations arose from the LSTM's computational intensity and the difficulty in interpreting its internal dynamics. While ARIMA lagged in performance, its simplicity and interpretability remained advantageous for understanding the structure of the time series.

The models' practicality was assessed by their ability to inform load balancing and energy procurement strategies, with LSTM providing the most robust framework for operational planning due to its accuracy.

Conclusion

The investigation concluded that while all models offer valuable insights, the LSTM network provided the most accurate forecasts for system demand. The applications of this work extend to energy sector operations, particularly in load forecasting and grid management.

For future work, I would explore hybrid models that combine the interpretability of ARIMA with the learning capabilities of LSTM networks. Additionally, incorporating exogenous variables like weather data and economic indicators could further enhance model performance.

# Results

# Analysis and discussion

# References

# Appendix