A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

INTEGRATED DATA-DRIVEN OPTIMISATION OF SUSTAINABLE ENERGY DISTRIBUTION NETWORKS IN PORTUGAL – ACCURACY OF ELECTRIC LOAD FORECASTING MODELS

ANA RITA SILVA

Work project carried out under the supervision of:

Sreyaa Guha

20/12/2023

Abstract

This thesis combines several research projects focusing on sustainable energy and utility services. It examines the relationship between network connectivity and utility service continuity, the adoption of Self-Consumption Small Production Unit for Self-Consumption (UPACs) in Portugal, the expansion of Electric Vehicle (EV) Charging Points Equipment (CPEs) in municipalities, and the effectiveness of electric load forecasting models. The findings reveal a complex relationship between network requests and service continuity, a significant influence of environmental factors on UPAC adoption, complex factors driving EV infrastructure, and the superior accuracy of *SARIMAX* over *XGBoost* in forecasting. These insights provide valuable guidance for Portugal's policymaking, infrastructure development, and energy management.

Keywords: Energy, Service Metrics, Remote Work, Electric Vehicles, Electric Vehicle Charging Points, Zero Emissions Resilient (ZER), Population Density, Charging Points Density, Electric Mobility, Grid Connections, Renewable Energy, Small Production Unit for Self-Consumption, Grid Development, Energy Distribution, *SARIMAX XGBoost*, *Load Forecast.*

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209

Table of Contents

List of Figures

List of Tables

Group Part

Abbreviations

ACEA	European Association of Automobile Manufacturers
ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller
AI	Artificial Intelligence
AIC	Akaike Information Criterion
APREN	Portuguese Renewable Energy Association
CPE	Charging Point Equipment
DGEG	General Direction of Energy and Geology
DSM	Demand-Side Management
DSO	Distribution System Operator
EDA	Exploratory Data Analysis
EESC	European Economic and Social Committee
END	Energy Not Delivered
EU	European Union
EV	Electric Vehicle
FEV	Full Electric Vehicles
Gw	Gigawatt
GWh	Gigawatt Hour
IEA	International Energy Agency
kV	Kilovolt
kW	Kilowatt
MAE	Mean Absolute Error
MAIFI	Momentary Average Interruption Frequency Index
MAPE	Mean Absolute Percentage Error

Group Part

ML	Machine Learning
MSE	Mean Squared Error
MW	Megawatts
OLS	Ordinary Least Squares Regression
PACF	Partial Autocorrelation Function
PEV	Plug-in Electric Vehicle
PV	Photovoltaic
REN	National Energy Network
RESP	Public Service Electric Network
RMSE	Root Mean Squared Error
RRP	Recovery and Resilience Plan
SAIDI	System Average Interruption Duration Index
SAIFI	System Average Interruption Frequency Index
SARIMAX	Seasonal Autoregressive Integrated Moving Average with Exogenous Factors
TIEPI	Total Energy Index of Power Interruptions
UPAC	Small Production Unit for Self-Consumption
XGBoost	Extreme Gradient Boosting
ZER	Zero Emissions Resilient

1. Introduction

This work project explores the complex aspects of utility services and sustainable energy in a world where technological advancement and sustainability concerns drive change. This collaborative effort brings together various research projects that explore different but related facets of energy systems and their significant influence on society.

The first stage in this investigation thoroughly examines how network connection requests affect utility service continuity. In an age where utility networks are growing increasingly complex and networked, it is critical to comprehend the relationship between these requests and service provision. This study explores how changing utility service management affects the energy industry and highlights its difficulties, such as increased remote work. The relationship between network connection requests and service continuity metrics offers insights into the demands on utility infrastructures, highlighting the critical need for resilient and adaptive systems.

In tandem with this, this work project investigates the critical role of renewable energy, concentrating on Portugal's UPACs. Adopting UPACs is a significant step toward improving sustainability and decentralizing energy production. This section of the study investigates the factors affecting how UPACs are implemented in Portuguese municipalities and provides a thorough understanding of the regional differences that influence the adoption of renewable energy by focusing on environmental, meteorological, and consumption influences. Prominent discoveries highlight the significance of meteorological factors, such as exposure to sunlight, which significantly influence the rate of UPAC installations.

The evolution of transportation through integrating Electric Vehicles (EVs) and related infrastructure is scrutinized. The research delves into the rise of Charging Points Equipment (CPEs) throughout Portuguese municipalities, marking an essential facet of mobility's progression. Moreover, the investigation delves into economic, demographic, and environmental influences on EV infrastructure growth. The analysis sheds light on the decisive role urban planning and policy play in sculpting CPE distribution and density, unraveling the complex interplay of these elements.

Continuing the exploration, the focus shifts to electric load forecasting, which is essential to contemporary power systems' functioning. Reliable load forecasting ensures that energy distribution networks can effectively supply demand. The study advances this field by evaluating the predictive power of sophisticated statistical and machine learning (ML) models applied to the Portuguese energy sector. We examine the efficacy of the eXtreme Gradient Boosting (*XGBoost*) and Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (*SARIMAX*) models and a hybrid approach that combines these techniques. This section of the work project incorporates meteorological factors, like temperature and precipitation, to improve the forecasting precision of energy consumption.

Various methodological techniques and reliable data sources support the group's investigations. Each segment of the thesis is informed by a tailored mix of analytical methods, including ML algorithms, *Granger* causality tests, and regression analyses, to thoroughly address the unique research questions and hypotheses within each section. The E-Redes Open Data Portal provides a wealth of data on utility services and energy metrics, establishing a solid empirical basis for the research. This deliberate and methodical data gathering, and analysis not only supports the study's relevance to real-world applications but also strengthens the results' trustworthiness.

The research uncovers complex dynamics within the utility sector, revealing a correlation between network connection requests and remote work order volumes, with minimal effects on service continuity indicators. It showcases the challenges service systems face with changing demands. Regarding renewable energy, the study emphasizes the pronounced impact of weather on the uptake of UPACs in Portugal, advocating for localized policy and planning.

EV CPE distribution in electric mobility varies significantly across municipalities, shaped by localized economic, demographic, and environmental factors.

Finally, exploring electric load forecasting highlighted the superior accuracy of *SARIMAX* models compared to *XGBoost*. Introducing a hybrid model showed additional improvements, signifying the advantages of combining different methodologies in predicting energy system demands.

The work project offers a thorough analysis of utility service management and sustainable energy systems, shedding light on the sector's prospects and obstacles by weaving together studies on load forecasting, renewable energy uptake, electric mobility infrastructure, and network connectivity. It aspires to enrich the field, providing insights for academia, industry, and policy formulation. The intent is to forge a knowledge base to inform policy crafting, infrastructure expansion, and service refinement, steering toward greater resilience, sustainability, and efficiency in the utility industry.

2. Literature Review

The costs associated with energy consumption, both financially and environmentally, are currently a major issue in our society. Renewable energy consumption was not previously a problem, but as society has grown more environmentally conscious, this idea is beginning to shift. The use and exploitation of renewable energies is now necessary due to the rise in energy demand brought on by the world's population growth and the fast expansion of emerging economies, which is also connected to the depletion of non-renewable resources. Considering the possibility that it will help avert an impending energy crisis and preserve Earth's ecology, we must keep funding its advancement.

Portugal has committed itself to the ambitious objective of attaining carbon neutrality by the year 2050, aligning with global and European targets established in implementing the Paris

Agreement. This commitment entails a substantial reduction of over 85% in greenhouse gas emissions, measured against the benchmark of 2005, coupled with the imperative of fostering a carbon sequestration capacity of approximately 13 million tons.

As the General Direction of Energy and Geology (DGEG) articulated, the anticipated energy transition in the coming decade necessitates an investment exceeding 25000 million euros. This undertaking requires a sophisticated coordination of collective wills, aligning policies, incentives, and financing mechanisms. To facilitate this transition effectively, the mobilization of a comprehensive suite of legal and planning instruments becomes imperative. These instruments serve as the strategic framework for achieving a tangible reduction in emissions while concurrently fostering investment, employment, and innovation. The DGEG emphasizes the crucial role of these measures in steering Portugal through the complexities of the energy transition and ensuring its success.

In pursuing comprehensive decarbonization, DGEG underscores its multifaceted significance as a strategic driver for investment and employment creation. The commitment to spearheading the energy transition is resolute, particularly through substantial investments in renewable production. As articulated by the DGEG, there is an imperative to more than double the installed capacity of renewable energy within the coming decade, with the ambitious goal of surpassing 80% of renewables in electricity production.

The outlined targets for Portugal by 2030 further emphasize the nation's dedication to sustainable practices. Aiming for 47% renewable energy in gross final energy consumption and a parallel target of 20% renewable energy in the transport sector by the specified timeline demonstrates a proactive stance toward achieving a cleaner and more sustainable energy landscape.

According to DGEG's perspective, the coming decade emerges as a critical juncture demanding heightened efforts to substantially curtail greenhouse gas emissions. This commitment entails the adoption of ambitious decarbonization targets, coupled with a deliberate focus on the integration of renewable energy sources and the enhancement of energy efficiency. The DGEG underscores the importance of ensuring a fair and inclusive transition, recognizing the social and economic dimensions as indispensable conditions for the success of this transformative vision.

To accelerate economic decarbonization, the XXIII Constitutional Government's program pledges various commitments, including:

- Developing a five-year carbon budget that specifies a multi-year horizon, expediting the execution of the 2030 National Energy and Climate Plans and the 2050 Carbon Neutrality Roadmap, supporting regional carbon neutrality roadmaps, providing methods for evaluating the impact of legislation on climate action, and eliminate administrative barriers that result in excessive context costs without environmental added benefit.
- Putting into practice the energy efficiency investment anticipated under the Recovery and Resilience Plan (RRP); 300 million euros for energy efficiency in residential structures, with a focus on lower-income families; 310 million euros for energy efficiency in the business sector and Public Administration service buildings; 715 million euros in the decarbonization of industry and 185 million euros in investment in hydrogen and renewable gases.
- Raise the capacity of solar energy production by at least 2 gw, keep holding auctions for new power plants, and encourage and assist the development of energy communities and self-consumption.
- Encourage green bond issuance, support micro-credit platforms for low-carbon investments, and facilitate collaboration between the Fund for Innovation, Technology,

and Circular Economy and the Environmental Fund to back decarbonization projects and enhance resource efficiency.

According to the IEA (International Energy Agency), in 2020, Portugal's electricity system was divided into two parts: a distribution system with high, medium, and low voltage lines and cables run by 13 distribution system operators (DSOs) and a transmission system with very-high voltage lines connected to Spain through nine cross-border interconnections. The transmission network has recently been enhanced to integrate new renewable generation. The national transmission network measured 9 036 km by the end of 2020, with 2 711 km operating at 400 kV, 3 780 km operating at 220 kV, and 2 545 km operating at 150 kV. The installed transformer capacity of the network was 38 463 megavolt-amperes.

Eleven of Portugal's thirteen DSO in 2020 were located on the country's mainland. 99.5% of customers in Portugal are connected at low voltage, and E-Redes is the only DSO for high-voltage and medium-voltage distribution systems. It also manages low-voltage distribution systems in 278 of the 308 municipalities in Portugal. The remaining 0.5% of consumers are served by ten more small-scale DSOs that run low-voltage distribution networks at the municipal level. Furthermore, the DSOs of the islands of Madeira and Azores are separate from the rest of the value chain.

To carry more renewable electricity to consumption centers, the transmission network has been strengthened in recent years due to the integration of high levels of new renewable generation. Additionally, the capacity of interconnection between Portugal and Spain has increased due to agreements made between the two nations to enhance the Iberian power market. Portuguese import capacity climbed from 1,112 MW to 2,970 MW in 2020, while export interconnection capacity went from 1,183 MW in 2010 to 2,925 MW in 2020. In Figure 1, it is possible to observe the National Transport Grid Map for the year 2020, as presented in the National Energy Network (REN).

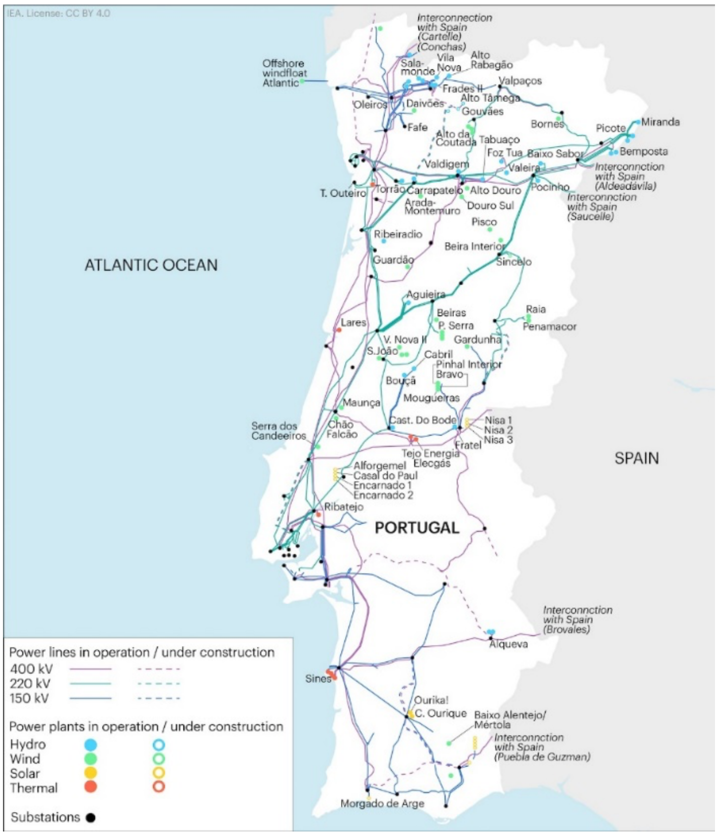


Figure 1. Portugal National Transport Grid Map. Source: REN.

According to the IEA, Portugal's electricity provision exhibits a division between renewable sources, predominantly wind and hydro, and fossil fuels, primarily natural gas and coal (Figure 2).

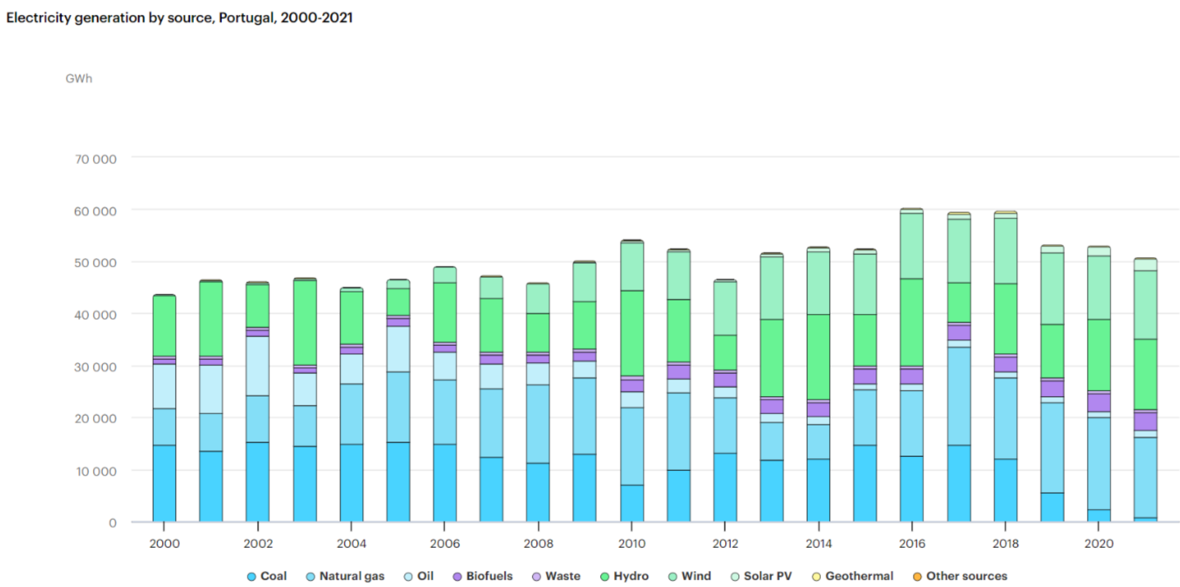


Figure 2. Electricity Generation by Source, Portugal, 2000-2021. Source: IEA.

In 2016 Portugal achieved a milestone by transitioning into a net electricity exporter, a trend sustained until 2019. Notably, this shift was attributed to the expansion of renewable energy generation. However, in the subsequent year, Portugal reverted to being a net importer of electricity (Figure 3).

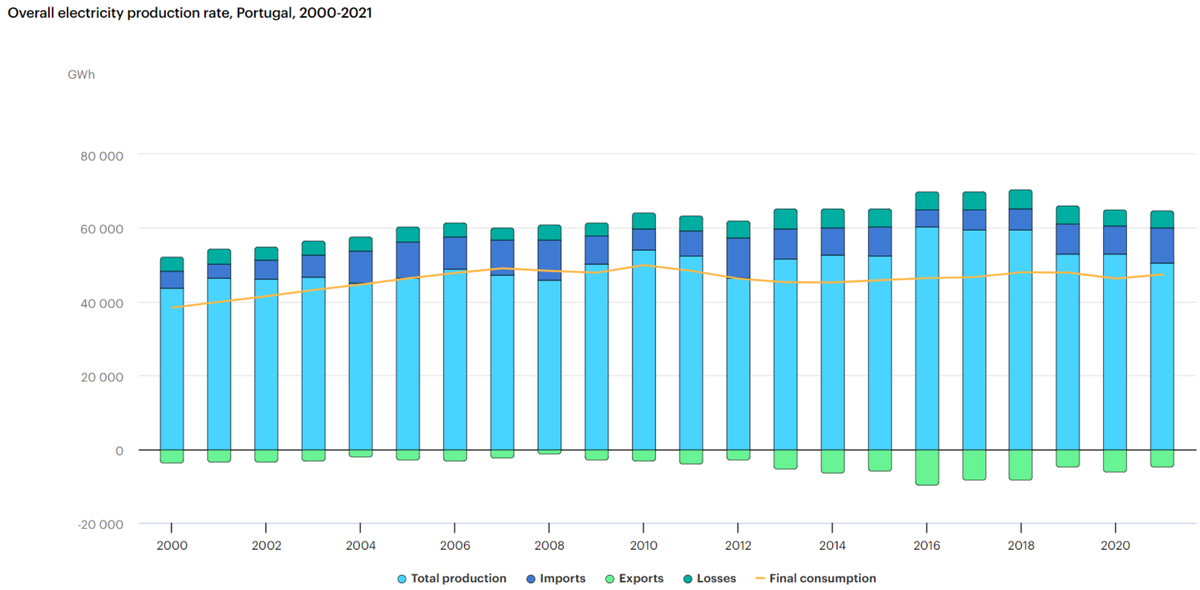


Figure 3. Overall Electricity Production Rate, Portugal, 2000-2021. Source: IEA.

In 2023, Portugal achieved a groundbreaking milestone in renewable energy production, setting a new record lasting almost a week. According to data from REN, from 4 AM on October 31st to 9 AM on November 6th, spanning 149 consecutive hours, Portugal produced 1102 gw per hour of renewable energy. This continuous output was more than sufficient to meet the combined industrial and residential consumption demands during the same period, totaling 840 GWh.

This accomplishment surpasses the previous record set in 2019, which stood at 131 hours of continuous renewable energy production. REN also highlights two additional milestones during this record-setting period. Between 10 AM on November 1st and 9 AM on November 5th, spanning 95 consecutive hours, Portugal's renewable production exceeded consumption without natural gas plants. During this time, Portugal even exported excess energy to Spain, surpassing the previous export record of 52 hours set in 2018.

Furthermore, from 10 PM on October 31st to 9 AM on November 6th, renewable energy production exceeded the national electrical system's needs, including the pumping requirements in hydroelectric reservoirs. Notably, hydroelectric production achieved these remarkable volumes significantly, showcasing Portugal's strides in sustainable and resilient energy systems.

Pragash, K, and P (2023) says that improved energy efficiency, less carbon emissions, better resilience and dependability, and enhanced system flexibility are some of the main attributes and advantages of smart grid technology. One major problem is integrating these variable energy sources into the current electrical infrastructure. The conventional electrical grid was intended to manage one-way power flows from the power plant to the end users. However, because renewable energy sources are spread geographically and produce power sporadically, they may cause grid oscillations and raise the possibility of blackouts. The following are a few of the major obstacles to incorporating renewable energy into the electrical grid:

- Intermittency and variability: The nature of renewable energy sources, like wind and solar power, is intermittent and variable, meaning that the amount of energy they produce varies according to the weather and other circumstances.
- Grid balancing and stability: The addition of significant amounts of intermittent renewable energy may impact these aspects of the system. To keep the system stable, grid managers must be able to swiftly balance supply and demand when the production of renewable energy sources varies.
- Economic challenges: Including the need for new price structures and incentives to promote the development of renewable energy projects, integrating renewable energy

sources into the grid may provide economic challenges. Legislators and regulators must get involved to establish a supportive regulatory framework for the growth and integration of renewable energy.

- Enhancing adaptability and responsiveness is essential to incorporate renewable energy sources into the grid successfully. The integration of renewables significantly influences grid operations and management, with critical impacts including:
- Lower carbon emissions: Using renewable energy sources like wind and solar can help lower carbon emissions and slow global warming. One of the main advantages of incorporating renewable energy into the electrical system is this.
- Intermittency and variability: The nature of renewable energy sources is frequently intermittent and variable, which means that the weather and other variables can affect how much energy they produce.

The transformative impact of data-driven techniques in energy distribution is paving the way for smarter, more efficient electricity networks. At the heart of this transformation are smart grids, which employ big data analytics to enable real-time monitoring and dynamic management of energy flows. This optimizes the distribution system's performance by utilizing data to forecast demand, integrate renewable sources effectively, and manage operations adaptively.

Ahmad et al. (2022) underscores the significance of advancements in ML technologies for optimizing decision-making within energy networks. These technologies are at the forefront of the shift towards a more integrated approach to energy distribution, driven by the urgency to address climate change and adhere to sustainability policies. The influence of ML is widespread, forging new paths in energy distribution across numerous areas, from the development of advanced energy materials to the strategic planning of energy systems, enhancing storage device capabilities, and bolstering energy efficiency.

Moreover, ML's reach extends beyond production. It plays a pivotal role in predictive maintenance, showcasing the extensive scope of ML in revolutionizing the sector. Data-driven methodologies are also redefining the management of energy outages. Big data analytics are indispensable for renewable energy and microgrid management, where precise and efficient power generation forecasting - rooted in extensive weather data - is critical. Such analytical prowess is essential for augmenting the efficacy of asset management and cooperative operations within the power industry, thereby enhancing grid reliability and stability.

Within the demand-side management (DSM) domain, big data analytics offers many insights extracted from large-scale electricity usage data, supporting the development of demand-side solutions and marketing plans. With historical load patterns, meteorological data, and societal trends contributing to a more nuanced understanding of future demands, load forecasting emerges as a critical research domain within intelligent grids.

Drawing insights from Crispim et al. (2014) in their paper on "Smart Grids in the EU with Smart Regulation: Experiences from the UK, Italy, and Portugal," the implementation of smart grids across the European Union (EU) offers a wealth of informative case studies. According to Crispim et al. (2014), the EU's adoption of smart grid solutions represents a collaborative effort towards cooperative development, as embodied in the EU 7th Framework Programme and Horizon 2020. These initiatives underscore the EU's commitment to integrating renewable energy sources, advancing demand response strategies, and optimizing the usage of electricity in new technologies like EVs and heat pumps.

Furthermore, the transition to smart grids in the EU, as expounded by Crispim et al. (2014), is influenced by several key factors. These include the urgency to lower greenhouse gas emissions, the challenges associated with managing grids that handle variable and unpredictable generations, and the integration of new categories of energy consumers, such as

EVs. This shift is also pivotal in redefining regulatory practices and shaping the market dynamics in various EU member states.

Building upon the insights from Crispim et al. (2014) regarding the implementation of smart grids across the EU, Portugal presents a case study of significant advancements in smart grid development. Soares et al. (2015) provide a detailed account of Portugal's major smart grid projects resulting from collaboration between academia and industry. These projects, namely InovGrid and REIVE, showcase the implementation of advanced metering systems, the integration of distributed energy resources such as EVs, microgeneration units, and energy storage devices into low-voltage networks, and the development of new functionalities for system operators based on microgrid concepts.

The InovGrid project aimed to develop an advanced metering infrastructure and functionalities that would enhance the intelligence of the distribution grid. This included managing and controlling microgeneration in low-voltage grids and providing conditions for customers to access new services. The project's real-world application was demonstrated in Évora, designated as InovCity, where the developed devices and control architecture were tested.

The REIVE project focused on the technical and market integration of renewables and EVs, emphasizing the role of EVs as flexible resources that can provide load flexibility and storage capacity. The project envisioned new grid-supporting functionalities and developed a market structure where an aggregator entity could provide market representation for the integrated EVs and microgeneration units.

Soares et al. (2015) also highlight the development of a Smart Grids and Electric Vehicle Laboratory as an outcome of the REIVE project. This laboratory is a testbed for evaluating smart grid concepts, control algorithms, communication solutions, and information exchange schemes. It has been instrumental in pre-validating functionalities intended for implementation in InovCity, including voltage and frequency regulation strategies and coordinated operation of microgeneration, storage, and EV charging.

This study also underlines the practical applications and benefits of smart grids. The InovGrid and REIVE projects not only facilitated the integration of renewable energy sources and EVs into the energy network but also demonstrated how smart grid technologies can enhance energy efficiency, promote consumer engagement in energy management, and contribute to the overall sustainability goals of the EU.

Remote work orders are usually issued in the energy distribution industry when system faults or disruptions need to be fixed. The consistency and dependability of the service provided can have a significant impact on how frequently these work orders are handled. By combining data-driven technologies and smart grids, service continuity can be improved, and the number of remote work orders can be decreased.

Big data analytics and ML enable smart grids to anticipate faults or disruptions and identify those areas in advance. With proactive maintenance and repairs made possible by this predictive capability, the number of remote work orders may be decreased. Using ML algorithms, historical data can be analyzed to spot trends or warning indications of impending failures, enabling prompt intervention to stop problems before they become more serious.

Furthermore, data-driven methods can enhance how remote work orders are handled when they arise. Through the analysis of data from multiple sources, including historical maintenance records, weather patterns, and real-time grid performance, utilities can more effectively allocate resources and prioritize work orders based on urgency. This can reduce the impact on consumers by resulting in quicker response times and more efficient issue resolution.

The European Economic and Social Committee (EESC) emphasizes the crucial symbiosis of energy self-production and self-consumption in the context of environmental protection, climate change mitigation, and addressing energy poverty. Unlike fossil fuels or nuclear energy, renewable technologies inherently possess a local dimension intimately tied to their geographical locations. Predominantly comprising photovoltaic (PV) solar and wind power, supplemented by small-scale hydropower, these technologies contribute to the revitalization of self-consumption in electricity, involving the direct utilization of locally generated energy.

The evolving landscape includes potentially incorporating a local variant of future green hydrogen. Over recent years, European and national legislation, particularly in select countries, has championed self-consumption. This support encompasses individual consumption through technologies like PV panels on rooftops and collective consumption facilitated by energy communities, local authorities, cooperatives, and similar entities deploying PV or wind farms. The clean and increasingly cost-effective energy produced, distributed, and consumed through these means holds significant promise for impacting accessibility and prices. Positioned as a valuable tool, it addresses energy poverty and contributes to the global effort against climate change.

Portugal reached a historic milestone in solar energy production in May 2022, becoming the fourth European nation to dramatically increase the percentage of renewable energy sources used in power production in the first five months of the year. Throughout the first half of 2022, the residential sector saw significant growth, suggesting a growing trend among families becoming more aware of the financial and environmental benefits of solar solutions. The use of government incentives contributes to this increase.

According to Mordor Intelligence, the need for an alternative electricity source and the desire to reduce the risk of climate change, in addition to the anticipated savings that are now more effective due to technical advancements, are the main reasons for this adoption by individuals.

The development of various technologies, such as EVs, is being encouraged by the decarbonization strategies devised by the EU and other European countries (Wee, Coffman, and La Croix 2018). Combined with low-carbon energy generation, they offer a viable way to lessen the transportation sector's substantial reliance on oil and gas products. Europe, along with China and the US, is one of the biggest EV markets in the world, with sales significantly growing year over year. However, due to various obstacles, such as increased investment costs, range anxiety, and a lack of adequate charging infrastructure, EVs still only make up a small portion of the global auto market in most countries.

The quantity of plug-in electric vehicles (PEVs) is impacted by government policy about PEVs. Three primary elements comprise government policies: incentives, regulations, and technology-related policies. Tax breaks and financial aid are two ways to encourage the purchase of PEVs. Numerous studies have shown a direct correlation between PEV sales and incentives. According to Sierzchula et al. (2014) the federal tax credit financed 30% of PEV sales.

The Portuguese government has been encouraging the development of electric mobility solutions and the use of Full Electric Vehicles (FEV). With the goal of reducing CO_2 emissions by 25% globally by 2020 in Portugal the Portuguese Electric Mobility Program (Mobi.E) had a partnership of enterprises and a network of municipalities to install and operate a pilot to promote electric mobility throughout the country. This initiative defined a market model that was piloted and developed in two later phases.

Portugal ranks among the top five European countries regarding charging points per 100 kilometers of road. According to ACEA (European Association of Automobile Manufacturers) research, Portugal ranks fourth on this statistic, trailing only the Netherlands, Luxembourg, and Germany. Furthermore, in May 2023, the share of EV sales was around 16%, compared to around 4.5% in Spain or Italy.

These are significant milestones, but they will need to be expanded to cover the entire region and the predicted growth in demand as the entire automotive industry is compelled to undergo an unprecedented electrical conversion to meet climate goals. Furthermore, the European Commission has set a maximum distance of 70 kilometers between two roadside charging locations.

The Portuguese problem stems mainly from more than 60% of homes needing a garage and relying only on the public network. On the other hand, the regional disparity in the distribution of this service has been identified as a weakness (Figure 4). According to *Mobi.E* data, in 2023, there will be 53 sockets for every 100 kilometers of highway in Portugal and 72 for every 100,000 residents.

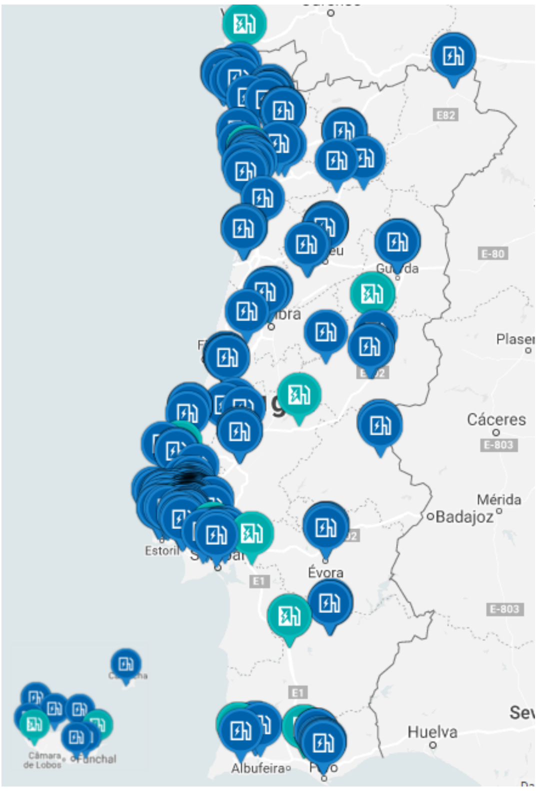


Figure 4. Distribution of CPEs in Portugal.

Electric load forecasting is pivotal in data-driven techniques for energy distribution optimization, ensuring efficient, reliable supply and supporting the intricate planning and operational demands of modern power systems. Accurate predictions of electricity demand are essential for allocating resources effectively and maintaining grid stability, particularly as grids become more complex and integrated with smart technologies.

Historically favored for their time series analysis, models like *ARIMA* have been instrumental in load forecasting by identifying and projecting historical load patterns into future trends. However, they often need to catch up in capturing the non-linear variability crucial for accommodating fluctuating renewable energy sources and evolving consumption patterns. The *SARIMAX* model addresses these limitations by incorporating exogenous factors, offering a more precise prediction of energy loads by considering influential variables like temperature, humidity, and consumer behavior.

The shift towards incorporating ML, exemplified by models like *XGBoost*, represents a leap in managing large-scale, non-linear data crucial for responding to the dynamic nature of contemporary energy demands. These models excel in processing vast quantities of data and uncovering complex relationships, a necessity for modern energy distribution systems that adapt to real-time changes and diverse energy inputs.

Integrating advanced forecasting models like *SARIMAX* and *XGBoost* into energy distribution systems is crucial for enhancing optimization strategies. Such integration leads to smarter grid management, allowing dynamic adjustments in energy supply based on accurate forecasts, improving efficiency, reducing waste, and better incorporating unpredictable renewable sources. Exploring these models, especially in hybrid forms, is critical to advancing energy distribution optimization, underpinning intelligent systems that ensure stable, sustainable supply amidst increasing demand and environmental sustainability goals. These techniques are fundamental in the data-driven optimization of energy systems, providing the essential intelligence for resilient and adaptable power infrastructures.

3. Discussion

This comprehensive analysis explores the intricate dynamics between executed network connection requests and their impact on service continuity indicators, remote work orders, and electric mobility inclusion. The findings indicate that increased network connection requests do not uniformly translate to improved service continuity indicators. Despite improvements in some indicators, others, such as the System Average Interruption Duration Index (SAIDI) and Energy not Delivered (END), show deterioration, highlighting a complex interplay of components. This complexity challenges existing models that predict a direct relationship between infrastructure investment and improved service outcomes.

The correlation between *network connection requests* and *end mt mwh* suggests that increased requests strain existing infrastructure or pose management challenges, affecting energy delivery. This observation has profound implications for utility companies, emphasizing the need to assess the quantity and quality of infrastructure investments.

Remote work orders data corroborates the importance of strategic planning in resource allocation and labor management. The increase in remote work orders with more network connection requests demonstrates a direct relationship between network expansion and the demand for associated services.

Simultaneously, the study delves into environmental and meteorological factors, particularly their influence on UPAC adoption rates. The positive coefficients for meteorological variables, highlighted by their feature importance in models like *XGBoost*, underscore the significance of environmental conditions like sunlight exposure. This finding is pivotal for policymakers, indicating that meteorological factors are crucial for accurately predicting UPAC adoption.

Comparing the Mean Squared Error (MSE) values across different supervised ML models in hypothesis testing reveals insights into their effectiveness in predicting UPAC adoption. The varying levels of predictive accuracy, especially the significant difference indicated by the *Ridge Regression* model, suggest complexities in the dataset that some models capture better than others. These insights are invaluable for developing and implementing policies promoting

UPAC adoption, with models like *Lasso Regression* offering robust frameworks for formulating targeted interventions.

The statistical analysis further examines the correlation between population density and the density of CPEs in Portuguese municipalities. Supported by both *Pearson* and *Spearman* Correlation analyses, a strong positive relationship is evident. However, the K-Means algorithm's clustering analysis reveals a nuanced picture: different clusters show varied relationships between population density, CPE density, and EV infrastructure development.

This analysis also challenges the assumption that higher income levels correlate with greater CPE density. *Regression analysis* demonstrates that population density is a significant predictor, but average income per city is not, contradicting previous hypotheses. Diagnostic tests reveal a more complex relationship than a linear model can capture, suggesting the need for a broader array of factors in examining EV infrastructure placement.

Additionally, the study examines the impact of Zero Emissions Resilient (ZER) urban policies. The correlation heatmap indicates a weaker correlation between ZER policy zones and CPE density, but further categorization into restricted and non-restricted zones reveals that restricted zones have higher CPE density. This supports the hypothesis that regulatory measures might encourage CPE installation.

Predictive modeling indicates that city size and population density account for about 50% of the variance in CPE density, hinting at unexplored variables that might influence this phenomenon. This moderate endorsement of the hypothesis suggests a complex interplay between urban policies, population density, and CPE infrastructure development.

The analysis also addresses the relationship between voltage level and electric mobility inclusion. *Cross-validation* confirms the significant association between these factors, indicating a genuine pattern rather than a sample-specific observation. *Logistic Regression* analysis further supports this, with models showing strong predictive capabilities and indicating a higher likelihood of electric mobility inclusion at low-voltage unique installations.

Model tuning with *GridSearchCV* optimizes the logistic regression model's hyperparameters, favoring a simpler model to mitigate overfitting risk. Sensitivity analysis confirms the model's robustness, suggesting that the findings are not influenced by the data proportion used for training or testing.

Lastly, the comparative analysis of forecasting models about *SARIMAX* and *XGBoost* models reveals that *SARIMAX* models demonstrate superior predictive capabilities. Lower MSE, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) values and higher R-Squared and custom accuracy rates support this. A hybrid approach combining *SARIMAX* and *XGBoost* offers a marginal improvement in forecast accuracy, confirming the effectiveness of a hybrid modeling approach in this context.

This study comprehensively explains the factors influencing service continuity, UPAC adoption rates, CPE infrastructure, and electric mobility inclusion. It highlights the complexity of these relationships and the importance of considering various factors for effective policy and infrastructure development in sustainable energy and technology adoption.

4. Limitations of the Study

In the limitations section of this thesis, several key constraints encountered during the research process need to be acknowledged, spanning various aspects of the study.

Primarily, the limitations are tied to the data and its scope. While the E-Redes statistics used in the study are extensive, they are confined to Portugal's electrical grid. This geographical and operational specificity implies that the findings might only be readily generalizable to other countries or energy sectors with further examination. Additionally, the reliance on secondary data introduces potential restrictions related to how the data was initially collected, including any inherent biases or inaccuracies in the data recording process. Given the rapidly evolving nature of the utility industry, driven by technological advancements and regulatory changes, the historical data may only partially capture the current dynamics of the sector.

Another limitation arises from the computational constraints that led to the use of district-level data despite having access to more detailed municipality-level data from E-Redes. This decision potentially limits the detail and insights that the forecasting models could have offered. Furthermore, if the data were a more extensive set of historical records, conducting a detailed municipality-by-municipality analysis would have enabled an understanding of specific local factors driving solar panel adoption. This approach could have provided more tailored and practical guidance to each municipality.

Moreover, the study faced challenges in exploring certain variables due to the need for comprehensive information and temporal data. These data availability issues hampered efforts to incorporate alternative data types like population density, CPEs locations, and income levels. The absence of longitudinal data mainly restricted the ability to perform a thorough temporal analysis, which would have been crucial for more profoundly understanding trends and patterns over time.

While the *Random Forest* model demonstrated superior predictive capability, it is a complex "black box" model that presents challenges in fully understanding its workings. Although significant, the insights derived regarding the importance of features should be approached with caution as they do not fully unravel the causal mechanisms at play.

In conclusion, the scope of this study's findings and conclusions is inherently influenced by these data limitations, affecting the generalizability and depth of the analysis. Future research in this domain would benefit significantly from access to broader datasets, including time-series data, to address these limitations and further build upon the foundational research presented in this thesis.

5. Conclusion

In concluding this research project, it is pivotal to revisit its core thesis: *"Integrated Data-Driven Optimization for Sustainable Energy Distribution Networks in Portugal"*. This central theme has guided an in-depth exploration into how data-driven strategies can revolutionize Portugal's energy distribution networks, emphasizing sustainability and efficiency. Through comprehensive analyses and discussions, the research has illuminated the critical role of technological advancements and innovative methodologies in shaping the future of energy distribution in a way that aligns with environmental sustainability and operational efficiency.

The project delved into several critical areas of Portugal's energy sector. It explored utility services, emphasizing the relationship between network demand and service continuity, revealing challenges in managing growing demands. In renewable energy, particularly UPACs, the study highlighted the crucial influence of environmental factors on adoption rates. The research on electric mobility infrastructure revealed the diverse factors influencing the distribution of EV Charging Points. Furthermore, in load forecasting, the project demonstrated the efficacy of *SARIMAX* models over others. These findings collectively underline the complexities in sustainable energy management, emphasizing the need for nuanced, data-driven policy-making and strategic infrastructure development in Portugal.

In the area of Service Continuity and Remote Services, the research highlighted the intricate relationship between infrastructure growth and service continuity, presenting complex challenges in utility service management. It underscored the need for utility companies to invest in expanding and improving their networks' resilience and efficiency. Additionally, the positive correlation between network connection requests and the increase in remote work orders reflected a responsive service sector adapting to changing operational paradigms and evolving customer needs. This part of the study calls for a reevaluation of current practices, suggesting a shift towards more adaptable and efficient utility service frameworks.

The research on Renewable Energy Integration Network revealed the critical role of environmental and meteorological factors in the adoption of UPACs in Portugal. The study's findings indicate the necessity for regionally specific renewable energy policies that consider local environmental conditions. This area of research offers a comprehensive understanding of the dynamics affecting UPAC adoption, providing valuable insights for policymakers and energy managers to propel sustainable energy strategies that align with Portugal's unique environmental landscape.

In the Electric Mobility and Grid Development section, the research provided robust evidence supporting the correlation between urban density and the distribution of EV CPEs in Portuguese municipalities. This study goes beyond simple correlations, shedding light on how economic, policy, and technical factors interplay in infrastructure deployment. It challenges traditional perspectives on income and policy incentives, underlining the complexity in developing EV infrastructure. This segment emphasizes the need for multifaceted urban planning and policy formulation to effectively deploy EV charging infrastructure within the sustainable transport landscape.

Finally in the Accuracy of Electric Load Forecasting Models section, the research assessed *SARIMAX* and *XGBoost* models' performance in predicting electric load, emphasizing the importance of weather variables and seasonal patterns. The *SARIMAX* model outperformed *XGBoost*, but a hybrid approach combining both models showed slight improvements in accuracy. These findings suggest the potential of integrated modeling methods in enhancing energy forecasts, crucial for optimizing sustainable energy distribution in Portugal. Further research in this area could explore detailed data usage and model refinement to advance forecasting efficacy.

Overall, the findings of this project cast a revealing light on Portugal's energy sector. The challenges in utility service scalability urge a rethinking of network management, crucial for service continuity. The influence of environmental factors on renewable energy adoption, particularly UPACs, underscores the need for policies attuned to regional environmental dynamics. In electric mobility, the diverse determinants of EV infrastructure deployment advocate for holistic urban planning approaches. The superiority of *SARIMAX* and hybrid models in forecasting signals a shift towards more accurate, data-driven energy management strategies. Together, these insights form a vital blueprint for developing a more sustainable and efficient energy landscape in Portugal.

Future research on *Integrated Data-Driven Optimization for Sustainable Energy Distribution Networks in Portugal* could explore several avenues. One area involves deepening the understanding of data integration techniques across different energy sources and distribution networks. Another research direction could focus on the application of advanced ML and AI (Artificial Intelligence) algorithms to predict and manage energy demand and supply more efficiently. Additionally, investigating the impact of smart grid technologies on enhancing network resilience and efficiency would be valuable. These studies could significantly contribute to optimizing Portugal's sustainable energy distribution, aligning with global sustainability goals.

2. Ana Rita Silva – Accuracy of Electric Load Forecasting Models

2.1 Introduction

Pursuing sustainable energy distribution networks is at the forefront of Portugal's national agenda as the country strides towards integrating innovative solutions for a greener, more efficient future. Central to this endeavour is optimizing energy systems through data-driven approaches, where accurate electric load forecasting becomes imperative. Electric load forecasting is a critical process in managing and operating power systems, ensuring that electricity supply matches demand with high efficiency and reliability.

This research aims to contribute to the broader objective of optimizing sustainable energy distribution networks in Portugal by evaluating and comparing the predictive performance of Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (*SARIMAX*) and eXtreme Gradient Boosting (*XGBoost*) models, as well as testing a hybrid approach that combines the strengths of both. The study focuses on the models' capabilities to incorporate weather variables and capture seasonal variations, which are crucial in accurately forecasting energy consumption within the Portuguese context. Addressing the research question, "*How does the accuracy of SARIMAX and XGBoost models, incorporating temperature, precipitation, and cloud cover, compare for time-series electric power load forecasting in Portugal?*" the comparative analysis revealed that *SARIMAX* models have superior predictive accuracy. The hybrid approach slightly enhances this accuracy, substantiating the potential of combined methodologies in this field.

The research is structured to provide an in-depth literature review, followed by a methodology section explaining the forecasting models, then the data analysis, results, discussion, and a conclusion with the implications for sustainable energy distribution and future research directions.

2.2 Literature Review

Electric load forecasting, critical for power system management, has traditionally relied on statistical models like *ARIMA*, which are effective for non-stationary time series data. With the advent of smart grid technologies, ML models such as *XGBoost* have emerged, offering scalability and efficiency in handling non-linear data. These models present distinct advantages: while *ARIMA* models are adept at identifying and projecting trends, *XGBoost* excels in processing complex data patterns, making them both valuable tools in the forecasting process.

Given these developments, a comparative analysis of *ARIMA* and *XGBoost* models in the context of electric power load forecasting becomes crucial to identify their pros and cons and to consider hybrid models that merge traditional statistical robustness with ML's advanced features.

The *SARIMA* model is widely recognized for its efficacy in forecasting time-series data, especially when seasonal patterns are prevalent. An extension of the *ARIMA* model, *SARIMA* is tailored to handle univariate time series data with a seasonal component, incorporating nonseasonal and seasonal features into its forecasting framework (Kramar and Alchakov 2023a).

SARIMA's application in electric load forecasting is precious due to its ability to model the trends and seasonality typically observed in electrical load data. These patterns often correlate with variables like population growth, industrial activities, and weather conditions, making *SARIMA* an ideal choice for capturing these fluctuations in electricity demand (Silfiani et al., 2023). However, the model's complexity and the requirement for stationarity in time series data necessitate rigorous preprocessing steps, such as the Augmented Dickey-Fuller (ADF) test, to ensure accurate predictions (Kramar and Alchakov 2023a).

The *SARIMAX* model, an advanced version of the *SARIMA* model, incorporates exogenous factors into its forecasting framework to enhance performance. These exogenous factors are external feature parameters that aid in reducing prediction errors, overcoming autocorrelation issues, and improving prediction results. In electric load forecasting, these exogenous factors are precious. They can integrate environmental and climatic variables to predict electricity demand fluctuations.

In electric load forecasting, *XGBoost* has proven to be particularly effective. The model's scalability allows it to process large-scale data efficiently (Ibukun 2019). Notably, *XGBoost* and similar regression-based models like Extra Trees Regressor have shown superior performance when modeling longer historical data, spanning from 2 to 6 years. Compared to autoregressive models like *SARIMA, XGBoost* recorded lower RMSE values for lengthier data, indicating its robustness in managing time-series fluctuations and extreme values. This effectiveness in handling large datasets and complex data patterns is a crucial strength of *XGBoost* in the field of electric load forecasting, as also supported by Wang (2019).

In stock market forecasting, Kumar et al. (2022) describe that the hybridization of the *SARIMA* and *XGBoost* models has shown remarkable effectiveness. This hybrid model leverages the strength of *SARIMA* in capturing seasonality aspects and the robustness of *XGBoost* in handling complex, non-linear patterns, especially in the volatile and unpredictable domain of stock prices. The approach considers historical price data and incorporates current market variables like open, high, low, and total volume, effectively optimizing for speed and performance. Notably, this hybrid model has demonstrated a high accuracy rate of 89.48%, complemented by favorable MAE and Mean Absolute Percentage Error (MAPE). Such performance metrics underscore the model's ability to provide reliable short-term and long-term stock value predictions.

Integrating weather variables, particularly temperature and humidity, is crucial in enhancing the accuracy of load forecasting models. Mohammadzadeh and Mosavi (2021) note that variables like temperature, humidity, and wind speed significantly influence the behavioral patterns of short-term electric load, demonstrating a high correlation with consumption rates. This study highlights how different seasons and external factors, such as school schedules and daylight hours, alter consumption patterns, affecting load forecasting. Similarly, Xie et al. (2018) underlines the importance of temperature, particularly dry bulb temperature, as a critical weather variable in load forecasting. The study observes that load increases in summer with rising temperatures due to cooling needs and in winter with decreasing temperatures to meet heating demands, illustrating the significant impact of these variables on electricity demand.

The literature review highlights the evolving landscape of electric load forecasting, where the precision of traditional statistical methods meets the adaptability of modern ML techniques. Despite the proven effectiveness of *SARIMA* in capturing seasonal variances and the robust nature of *XGBoost* in handling intricate data patterns, there remains a need for a comprehensive comparison within the unique context of Portugal's energy sector. Moreover, the literature suggests the potential for hybrid models to capitalize on the complementary strengths of both methods. Within this intersection of established practices and innovative approaches, this research situates its inquiry, leading to the formulation of a critical research question and corresponding hypotheses. The ensuing segments will delve into the empirical evaluation of these models, thereby contributing to the strategic goal of enhancing the reliability and efficiency of Portugal's electric power grid.

2.3 Research Question and Hypothesis

In addressing the challenge of sustainable energy distribution in Portugal, it is essential to answer the following research question: "*How does the accuracy of SARIMAX and XGBoost* *models, incorporating temperature, precipitation, and cloud cover, compare for time-series electric power load forecasting in Portugal?"*. This inquiry stems from thoroughly examining the current landscape in electric load forecasting, acknowledging the transition from traditional statistical approaches to advanced ML techniques. The literature reveals a dichotomy between *SARIMA*, known for its proficiency in handling seasonal data within non-stationary time series, and *XGBoost*, celebrated for its efficiency in large-scale data processing and non-linear pattern recognition. Given the intricate interplay of weather variables like temperature, precipitation, and cloud cover in electric load forecasting, this research seeks to evaluate and compare the predictive capabilities of these two distinct yet complementary modeling approaches within the Portuguese context.

The hypotheses formulated for this research are as follows:

1. The primary hypothesis posits that "*SARIMAX models will demonstrate a higher predictive capability than XGBoost models."*
2. A secondary hypothesis suggests that "*The hybrid approach combining SARIMAX models with XGBoost algorithms will outperform standalone SARIMAX models in forecasting electric power load."*

These hypotheses are anchored in the empirical evidence and theoretical foundations presented in the literature. Hypothesis 1 is based on the proven ability of the *SARIMAX* model to capture seasonal variations in electric load data and its integration of exogenous factors. This hypothesis is also motivated by the solid and data-driven attributes of *XGBoost*, which is particularly effective in managing complex, non-linear data structures. Additionally, the successes of hybrid models in other forecasting domains have inspired the exploration of a combined *SARIMAX-XGBoost* approach in Hypothesis 2. This proposed hybrid model aims to merge *SARIMAX's* seasonal forecasting aptitude with *XGBoost's* capacity for intricate, non-linear data patterns, potentially elevating the precision and effectiveness of electric power load forecasting in Portugal.

2.4 Methods

2.4.1 Research Design and Model Selection

The research design of this study is structured as a quantitative comparative analysis aimed at evaluating the predictive accuracies of two distinct forecasting models: *SARIMAX* and *XGBoost*. The study employs *SARIMAX* and XGBoost models for their recognized efficiency in forecasting. *SARIMAX* was chosen for its ability to handle seasonal trends in electric load data, while *XGBoost* was utilized for its strength in analyzing large, complex datasets. This research design was selected to rigorously assess the forecasting performance of these models under varying conditions influenced by weather variables, such as temperature, precipitation, and cloud cover, which are critical to electric power load forecasting in Portugal.

A suite of statistical tests and error metrics will be employed to compare the *SARIMAX* and *XGBoost* models. These include the MSE, RMSE, MAE, MAPE, and the coefficient of determination, R-Squared. These metrics were chosen due to their extensive use and acceptance in statistical and machine-learning fields for quantifying prediction accuracy. The development and tuning of each model were executed with precision, ensuring that the *SARIMAX* model's parameters optimally capture the seasonal and non-seasonal patterns. In contrast, the *XGBoost* model's configuration is meticulously adjusted to avoid overfitting and maximize forecast accuracy. In addition to these standard measures, the analysis will incorporate custom accuracy metrics that evaluate the percentage of predictions that fall within 10%, 5%, and 1% tolerance levels, providing a nuanced evaluation of the models' performance in practical scenarios.

The analytical techniques selected for this study are particularly suitable for the research question, as they enable a comprehensive comparison between the models' capabilities to forecast electric power load accurately. By evaluating the models across these diverse metrics, the research aims to determine which model performs better and explore the potential of a hybrid *SARIMAX-XGBoost* model. This dual-level analysis, incorporating traditional statistical measures and tailored accuracy evaluations, aligns precisely with the intent to optimize sustainable energy distribution networks through improved forecasting methodologies.

2.4.2 Data Selection

The selection of robust datasets critical for the accurate modeling and forecasting of electric power load in Portugal underpins the empirical foundation of this study. The primary dataset is sourced from E-Redes, the Portuguese electricity DSO, which provides granular, hourly consumption data categorized by postal code. This dataset is pivotal due to its comprehensive coverage and high temporal resolution, allowing for a detailed analysis of consumption patterns across different regions and times. Complementing this, to integrate the significant impact of weather conditions on electricity consumption, hourly weather data, including temperature, precipitation, and cloud cover, has been procured from Open-Meteo API. These weather variables correlate strongly with power load, and their inclusion is expected to enhance the predictive accuracy of the models. Together, these datasets form a comprehensive empirical basis for developing a forecasting model that reflects the complex interplay between human activity, geography, and weather patterns in the context of Portugal's energy distribution network.

2.5 Data Analysis

2.5.1 Data Description

The dataset for this study consists of hourly electric power load data, weather variables (*temperature, precipitation,* and *cloud cover*), and geographical context (*district* and its *elevation*). This multi-source dataset spans from October 1, 2022, to August 31, 2023, and encompasses a comprehensive structure of 144720 rows and seven columns, representing hourly data points across multiple districts.

The variables included in the dataset are as follows:

1. ***datetime***: An hourly timestamp marking each observation, formatted as "YYYY-MM-DD HH:MM: SS."
2. ***district***: The name of the district associated with each observation provides a geographical identifier for the energy consumption data.
3. ***active_energy_kWh***: The amount of active energy consumed during each hour, measured in kilowatt-hours, indicative of the electrical demand in the corresponding district.
4. ***temperature_2m (°C)****:* The air temperature measured two meters above the ground, an environmental variable that can significantly affect energy usage patterns.
5. ***precipitation (mm)****:* The volume of precipitation measured in millimeters, captured hourly, providing insights into the weather conditions at the time.
6. ***cloudcover (%)****:* The percentage of the sky obscured by clouds.
7. ***elevation****:* The elevation in meters of the district.

With its granular hourly resolution for each district within the specified time range, the dataset is designed to facilitate a detailed time-series analysis. This analysis will incorporate temporal patterns in energy consumption and examine the influence of climatic and topographical factors.

2.5.2 Exploratory Data Analysis

Before delving into the modeling phase, an EDA was conducted to visualize the data's distribution, identify outliers, and detect underlying patterns or trends.

Figure 5 illustrates significant variations in average active energy across different districts in Portugal, with coastal districts exhibiting markedly higher consumption levels than inland districts.

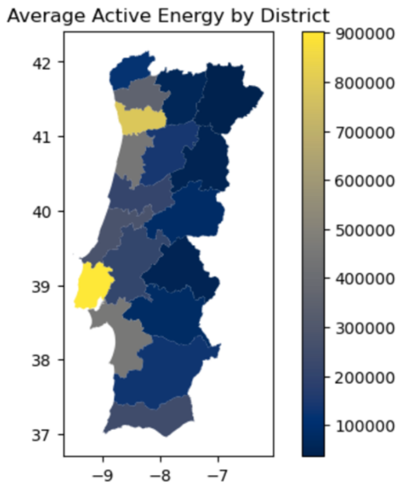


Figure 5. Average Active Energy by District.

Appendix A presents a series of time series plots, each illustrating the hourly active energy for one month across the different districts in Portugal. These graphs reveal that the intrinsic seasonal characteristics of electric power load are evident across all districts, although at varying magnitudes.

The variation in energy use across districts, as shown in Figure 5, and the clear seasonal trends in Appendix A's time series plots underline the necessity of integrating region-specific analysis for energy demand forecasting.

Figure 6 illustrates the relationship between weather variables - temperature, precipitation, cloud cover - and active energy. From the temperature plot, one might infer a potential relationship between temperature and energy usage, as indicated by a possible clustering or pattern. The relationship appears less clear for precipitation and cloud cover, with data points spread out, suggesting a weaker or more complex interaction with energy consumption. These plots lead to the hypothesis that weather conditions have varying degrees of impact on energy consumption, with temperature potentially being a more significant predictor than precipitation or cloud cover. However, the scattered and diffuse nature of the

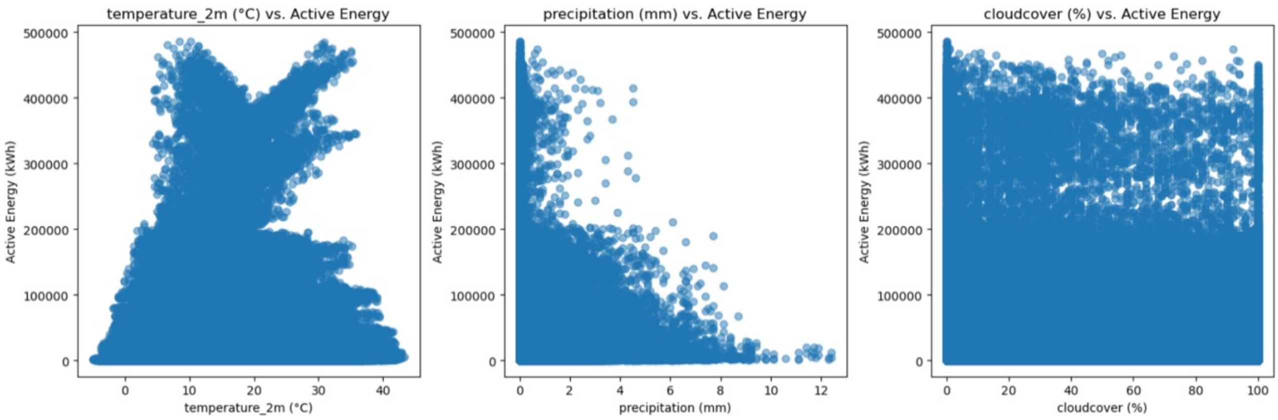


Figure 6. Scatter Plots of Weather Variables vs. Active Energy.

points, particularly in the latter two plots, indicates that a simple linear relationship may not exist, and more sophisticated modeling techniques could be required to understand these dynamics fully.

Outliers within the active energy variable were identified and handled using the *Z-score* method, a standard statistical approach for detecting anomalies in univariate data. Following the guidance of Iglewicz and Hoaglin (1993), data points exceeding three standard deviations from the mean were deemed outliers. Upon identification, these outliers were addressed using the forward-fill method, which imputes the missing or anomalous values with the subsequent valid value. This technique ensures that the temporal sequence of the data remains intact, avoiding the exclusion of any timestamps that could otherwise introduce bias into the time-series modeling process.

In the feature engineering phase, lagged features were introduced to capture temporal dependencies in energy consumption. Specifically, lagged values of active energy were created for one- and two-time steps back (*lag_1* and *lag_2*) within each district.

Additionally, a categorical feature *day_type* was engineered to differentiate between weekdays, weekends, and holidays, recognizing the potential impact of these different periods on energy consumption patterns. The day type feature was then one-hot encoded, with the first category dropped to prevent multicollinearity. These newly crafted features enrich the dataset, allowing the forecasting models to leverage temporal and calendar-based nuances.

2.5.3 *SARIMAX* Model

To address the research question of comparing the accuracy of *SARIMAX* and *XGBoost* models for time-series electric power load forecasting in Portugal, an initial step involved determining appropriate parameters for the *SARIMAX* model, which is characterized by non-seasonal parameters (p, d, q) and seasonal parameters (P, D, Q, s). Given the extensive computational resources required to tailor parameters for each district individually, a consolidated approach was taken. The dataset was aggregated to provide a single time series encapsulating the active energy across all districts. This aggregated dataset was then used to identify parameters that could be generalized to each district-specific model.

In preparation for modeling, the stationarity of the aggregated electric power load data was assessed using the ADF test. This rigorous statistical test determines the presence of unit roots (Dickey and Fuller 1979). The ADF test yielded a statistic of -10.07 and an extremely low p-value of approximately 1.25×10^ (-17), decisively falling below the standard alpha level of 0.05. These results confirm the stationarity of the series and suggest that a differencing order of d = 0 is very likely, validating its use in its current form for *SARIMAX* modeling and underpinning the model's suitability for forecasting electric power load across Portugal's districts.

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots in Appendix B are pivotal tools in identifying the parameters for the model (Box et al.

1994). The ACF plot identifies a clear seasonal pattern with significant autocorrelations at the 24th lag, which aligns with the 24-hour cycle in the data, representing the inherent daily seasonality of electricity consumption. This consistent pattern supports the selection of s=24 for the seasonal component of the *SARIMAX* model. The PACF plot exhibits a notable initial decline and several subsequent significant spikes, suggesting a more complex autoregressive structure. A comprehensive grid search approach was then executed to determine the optimal *SARIMAX* model parameters, focusing on a range of potential non-seasonal and seasonal values with a daily cycle of 24 hours. The resulting best-performing model was identified as *SARIMAX* (0, 0, 0) x (1, 1, 1, 24), indicating that the seasonal elements alone were adequate for modelling, capturing the cyclical daily patterns in energy usage, as confirmed by the Akaike Information Criterion (AIC) measure.

To maintain the granularity of the data and capture the unique energy consumption patterns across different districts, the models were individually trained for each district. This approach ensures that the nuances of local energy usage, influenced by varying factors such as weather, geography, and human activity, are reflected accurately in the forecasts. For a comprehensive understanding of the models' performance, the appendices provide detailed summaries of each district's *SARIMAX* model (see Appendix C). These summaries contain the estimations of coefficients, their standard errors, and diagnostic tests. Additionally, visual comparisons between the actual and predicted energy consumption values for each district are also presented in the appendices (see Appendix D), demonstrating the accuracy of the forecasts generated.

The *SARIMAX* models' performance was evaluated comprehensively, unifying the predicted values and the actual active energy data from each district to analyze the model as an integrated whole. This evaluation method allows for an assessment that reflects the overall predictive accuracy across all districts, ensuring that any district's performance does not skew the results. Table 15 provides the metrics values used to quantify the forecasting accuracy.

Table 1. SARIMA's Evaluation Metrics

Evaluation Metric	
MSE	65137263.6886
RMSE	8070.766
MAE	4221.1842
MAPE	1.67%
R-Squared	0.9989
Custom accuracy (10%)	99.51%
Custom accuracy (5%)	96.59%
Custom accuracy (1%)	41.70%

As shown in the table, the MSE and the RMSE are particularly elevated, which is indicative of the significant variances within the dataset, potentially influenced by fluctuations in energy consumption. Despite this, the MAE suggests that, on average, the model's predictions are close to the actual values. The MAPE at 1.67% indicates a high level of accuracy in percentage terms, and the R-Squared value of 0.9989 confirms the model's effectiveness in capturing the variance of the active energy.

Furthermore, custom accuracy metrics further demonstrate the model's precision, with most predictions falling within a 10% margin of error and a substantial proportion within the tighter thresholds of 5% and 1%. Additionally, Appendix E presents the scatter plot of residuals, displaying a random dispersion of data points around the zero line, indicating no apparent systematic error in the model's predictions. These results underscore the model's capability to produce forecasts with a high degree of reliability for practical applications in electric power load forecasting.

2.5.4 *XGBoost* Model

For the *XGBoost* model, the approach differed from the *SARIMAX* model as there was no requirement to build separate models for each district. Instead, the district variable was one-hot encoded, effectively transforming it into a set of binary features to capture the distinct characteristics of each district within a single model. The datetime index was also decomposed into separate features—*year, month, day,* and *hour*—to provide the model with a granular temporal context, which is particularly important for capturing patterns that may vary across different timescales.

The hyperparameter tuning process for the *XGBoost* model was accomplished through grid search and cross-validation, an exhaustive search technique that iterates over a predefined set of hyperparameters to find the combination that yields the best performance based on a scoring metric—in this case, the RMSE. This method is beneficial for models like *XGBoost*, where the interaction of different parameters can significantly influence the final model's accuracy.

The *XGBoost* model's grid search focused on key hyperparameters: *tree depth, minimum child weight, gamma, subsample, colsample by tree, and learning rate, balancing model complexity* and *regularization* to prevent overfitting. This approach identified the most effective hyperparameters for the *XGBoost* model, aiming to minimize the RMSE and enhance the predictive accuracy of the electric power load forecasts. The optimal parameters identified and implemented for the model training are shown in Table 16.

Table 2. XGBoost Parameters.

Parameters			
colsample_bytree	1.0	*min_child_weight*	10
gamma	0.5	*subsample*	0.6
learning_rate	0.3	*eta*	0.3
max_depth	4		

The model was trained to employ the optimal parameters identified, with the training process spanning 100 iterations. For a detailed insight into the model's internals, refer to Appendix F, which presents a table illustrating the weight and importance of each feature.

Table 3. XGBoost's Evaluation Metrics.

Evaluation Metrics	
MSE	92549493.2655
RMSE	9620.2647
MAE	5482.603
MAPE	2.59%
R-Squared	0.9985
Custom accuracy (10%)	98.32%
Custom accuracy (5%)	87.30%
Custom accuracy (1%)	28.02%

The performance metrics of the *XGBoost* model in Table 17 present a nuanced view of its forecasting accuracy for electric power load. The MSE and RMSE are considerably high at approximately 92.5 million and 9620, respectively, indicating significant dataset variances. The MAE stands at 5482.6, which is significant but mitigated by the model's high R-Squared value of 0.9985. This suggests that despite significant individual errors, the model's predictions generally agree with the actual values. For a detailed comparison of actual versus predicted active energy by the XGBoost model, see Appendix G.

The results in the custom accuracy measures are promising, which provide an idea of the model's precision within specified tolerance levels. With 98.32% of predictions falling within 10% of actual values and 87.30% within a 5% margin, the model demonstrates high accuracy in its forecasts. However, the accuracy drops to 28.02% when the margin is tightened to 1%, underscoring the model's challenge in predicting near-perfect precision. This could be a factor for applications requiring exact load forecasting. Furthermore, there appear to be no systematic errors in the model's predictions, as evidenced by the scatter plot of residuals in Appendix H, which shows a random distribution of data points around the zero line. Overall, the model appears robust for forecasting electric power load, with a performance that suggests a balance between capturing general trends and acknowledging specific fluctuations.

2.5.5 Combination of *SARIMAX* and *XGBoost*

An innovative approach was adopted to achieve enhanced forecasting accuracy by combining the predictive powers of *SARIMAX* and *XGBoost* models. This involved the creation of two distinct dataframes—one with the *XGBoost* predictions along with district identifiers and the corresponding datetime indices from the test set and the other containing the *SARIMAX* predictions with similar district information and datetime indexing.

These dataframes were meticulously merged to ensure a seamless alignment of predictions from both models for each district and timestamp. Subsequently, this merged dataframe was integrated with the test set to append the actual observed values alongside the predicted ones. The study used a straightforward yet efficient method for creating a combined forecast to combine the predictive powers of the *SARIMAX* and *XGBoost* models. To do this, the corresponding forecasts from each model for the exact date and district were averaged. The idea behind such an averaging strategy is that aggregating forecasts can lower the variance of the prediction errors, which could result in a more stable and probably more accurate forecast. This approach assumes that the models capture various parts of the data separately and that combining these diverse insights would improve the averaged forecasts.

This method leverages the strengths of both models—*SARIMAX*'s capacity to handle seasonal variations and *XGBoost's* prowess in capturing complex nonlinear relationships—resulting in a robust prediction that is expected to be more accurate and reliable than either model individually. This composite forecast is a strategic blend expected to mitigate individual model biases and offer a well-rounded forecast. To assess the precision and effectiveness of the integrated *SARIMAX* and *XGBoost* approach, the same set of evaluation metrics was applied, with the detailed results displayed in Table 18:

Table 4. Hybrid Model Evaluation Metrics.

Evaluation Metrics	
MSE	62219419.9608
RMSE	7887.9287
MAE	4166.45
MAPE	1.76%
R-Squared	0.999
Custom accuracy (10%)	99.53%
Custom accuracy (5%)	96.18%
Custom accuracy (1%)	39.00%

Analyzing the evaluation metrics, the MSE and RMSE are relatively low, indicating good model accuracy with limited prediction errors. The MAE and MAPE are also low, suggesting precise predictions with minimal deviation from actual values. An R-Squared value close to 1 reflects the model's excellent fit to the data. Furthermore, custom accuracy metrics demonstrate high reliability, with nearly all predictions within a 10% margin of error and a significant majority within a 5% range, indicating the model's robustness and potential utility in practical applications. However, the custom accuracy at a 1% tolerance is 39.00%, highlighting that while the model is robust for most practical applications, there is a potential for further refinement to meet the stringent demands of scenarios where ultra-precise predictions are crucial.

2.6 Discussion

The comparative analysis of the forecasting models about the set hypotheses reveals insightful outcomes. The first hypothesis posited that *SARIMAX* models would exhibit superior predictive capabilities compared to *XGBoost* models. Evaluating the metrics, *SARIMAX* models demonstrated lower MSE, RMSE, and MAE values and slightly higher R-Squared and custom accuracy rates at all levels, supporting the hypothesis. Specifically, *SARIMAX*'s R-Squared value of 0.9989 and higher accuracies across 10%, 5%, and 1% thresholds, compared to *XGBoost's* 0.9985 and lower accuracies, highlight its enhanced performance.

Turning to the second hypothesis, which anticipated that a hybrid approach combining *SARIMAX* and *XGBoost* would outperform the standalone *SARIMAX* model, the results also confirm this prediction. The hybrid model yielded an improved MSE and an R-Squared value of 0.999, indicating a marginally better fit than the standalone *SARIMAX* model. Furthermore, the combined model's custom accuracy at a 10% tolerance level reached 99.53%, surpassing the *SARIMAX* model's 99.51%. However, it is worth noting that the performance improvement is modest, with the hybrid model offering only a slight enhancement over the standalone *SARIMAX* model.

In conclusion, both hypotheses are confirmed. The *SARIMAX* model proves to be more effective than the *XGBoost* model when used in isolation for electric power load forecasting. Furthermore, combining *SARIMAX* with *XGBoost* leads to a slight improvement in forecast accuracy, validating the effectiveness of a hybrid modelling approach in this context.

2.6.1. Limitations of the Study

This study's scope was limited to district-level data due to computational constraints despite having access to more granular municipality-level data from E-Redes. This choice may affect the detail and potential insights of the forecasting models.

2.7. Conclusion

In this study, we sought to bolster the sustainability of Portugal's energy distribution networks by evaluating the predictive capabilities of *SARIMAX* and *XGBoost* models, especially in their ability to incorporate crucial weather variables and seasonal patterns for accurate electric load forecasting. Our findings indicate that the *SARIMAX* model excels over *XGBoost* in predictive accuracy. However, a synergistic hybrid model combining *SARIMAX* with *XGBoost* demonstrates a slight but meaningful enhancement in performance.

This supports the potential of integrated modeling methods in improving energy system forecasts. These insights contribute to optimizing Portugal's sustainable energy distribution, highlighting the pivotal role of advanced predictive models in meeting the nation's energy objectives. Future research should consider incorporating more detailed data and alternative modeling techniques to refine forecast accuracy and identify specific conditions under which *XGBoost* models may surpass *SARIMAX* models in forecasting efficacy. Understanding the triggers and thresholds for *XGBoost's* superior performance could further enrich our modeling toolkit, ultimately aiding the global pursuit of energy sustainability.

References

Abdulqader, Hussein, and R. Swaroop. "LOAD FORECASTING FOR POWER SYSTEM PLANNING AND OPERATION USING ARTIFICIAL NEURAL NETWORK AT AL BATINAH REGION OMAN." *DOAJ (DOAJ: Directory of Open Access Journals)*, August 1, 2012. https://doaj.org/article/93ca549faa004cf89749f299ddd5bc7f.

Aguiar, Carla. 2022. "Portugueses rendem-se à energia solar em plena crise energética." *Diário De Notícias*, August 8, 2022. https://www.dn.pt/dinheiro/portugueses-rendem-se-a-energia-solar-em-plena-crise-energetica-15078038.html.

Ahmad, Tanveer, Rafał Madoński, Dongdong Zhang, Chao Huang, and Asad Mujeeb. "Data-Driven Probabilistic Machine Learning in Sustainable Smart Energy/Smart Energy Systems: Key Developments, Challenges, and Future Research Opportunities in the Context of Smart Grid Paradigm." *Renewable & Sustainable Energy Reviews* 160 (May 1, 2022): 112128. https://doi.org/10.1016/j.rser.2022.112128.

Alharbi, Fahad Radhi, and Dénes Csala. "A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach." *Inventions* 7, no. 4 (October 16, 2022): 94. https://doi.org/10.3390/inventions7040094.

Chandra, Ambarish, Sumeet Gulati, and Milind Kandlikar. 2010. "Green Drivers or Free Riders? An Analysis of Tax Rebates for Hybrid Vehicles." *Journal of Environmental Economics and Management* 60 (2): 78–93. https://doi.org/10.1016/j.jeem.2010.04.003.

Crispim, João, José Braz, Rui Castro, and José Esteves. "Smart Grids in the EU with Smart Regulation: Experiences from the UK, Italy and Portugal." *Utilities Policy* 31 (December 1, 2014): 85–93. https://doi.org/10.1016/j.jup.2014.09.006.

Group Part"

Da Fonseca Monteiro Pereira, Rita Manuela. "Redes de Distribuição de Energia Elétrica de Média e Baixa Tensão : Estágio Na Helenos, S.A," June 27, 2018. https://comum.rcaap.pt/handle/10400.26/25475.

Dickey, David A., and Wayne A. Fuller. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." Journal of the American Statistical Association 74 *Www.Jstor.Org*. https://www.jstor.org/stable/2286348.

European Economic and Social Committee. 2023. "Energy – Producing and consuming own energy key to fighting climate change and energy poverty." November 22, 2023. https://www.eesc.europa.eu/pt/news-media/news/energy-producing-and-consuming-own-energy-key-fighting-climate-change-and-energy-poverty#related-content.

Gouveia, Clara, David Rua, Filipe Soares, Carlos Moreira, Pedro Matos, and João Lopes. "Development and Implementation of Portuguese Smart Distribution System." *Electric Power Systems Research* 120 (March 1, 2015): 150–62. https://doi.org/10.1016/j.epsr.2014.06.004.

Ibukun, Gabriel Dada. "Analysis of Electric Load Forecasts Using Machine Learning Techniques." **MSc Research Project, National College of Ireland, 2019.**

IEEE Conference Publication | IEEE Xplore. "Analysis and Prediction of Stock Price Using Hybridization of SARIMA and XGBOOST," March 10, 2022. https://ieeexplore.ieee.org/abstract/document/9767868?casa_token=j_lk7uZKBQIAAAAA:-DN5zwgcbqCy4X5WfUi6IGQ1WV3FzohpiuI6ENRy6ucvP3EObToy-Fx-t6NjnLZTBzUPNQ9Mwtw.

IEEE Conference Publication | IEEE Xplore. "Comparing Various Combined Techniques at Seasonal Autoregressive Integrated Moving Average (SARIMA) for Electrical Load Forecasting," July 26, 2023. https://ieeexplore.ieee.org/abstract/document/10221130?casa_token=ONErvE5NUgAAAAA:a2kewIOc8Dr51E1jlyNgW3hUlzBC3jL2iswkci4_60y4Bo_07GCLJT-pIeJ6CAIlvYxZiwjnHc4.

IEEE Conference Publication | IEEE Xplore. "Evaluation of the Machine Learning Techniques for Forecasting the Seasonal Time Series," May 15, 2023. https://ieeexplore.ieee.org/abstract/document/10158869?casa_token=pon9Ieoh5wEAAAAA:69gBKHFRy3HuKo2qECZE3PimDms4lWNiG2da5EE5aJCU3-5gqXjCFKWLI4Zfr8Fhavri8N5hAsA.

IEEE Journals & Magazine | IEEE Xplore. "An Experimental Machine Learning Approach for Mid-Term Energy Demand Forecasting," 2022. https://ieeexplore.ieee.org/document/9945969?denied=.

IEEE Journals & Magazine | IEEE Xplore. "Relative Humidity for Load Forecasting Models," January 1, 2018. https://ieeexplore.ieee.org/abstract/document/7448451?casa_token=Z2pUTpUDCjMAAAAA:CvD802USMqoxZIon76WFzlHvzPBaelASVueYoSMVQ8RRxgOdiokCLAetM9zBZmcIwE0BRSJavz8.

Iglewicz, Boris, and David Hoaglin. 1993. "Volume 16: How to Detect and Handle Outliers." In The ASQC Basic References in Quality Control: Statistical Techniques, edited by Edward F. Mykytka. Milwaukee: American Society for Quality Control. https://books.google.pt/books?hl=en&lr=&id=FuuiEAAAQBAJ&oi=fnd&pg=PP1&dq=Iglewicz,+Boris,+and+David+Hoaglin.+1993.+%22Volume+16:+How+to+Detect+and+Handle+Outliers.%22++In+The+ASQC+Basic+References+in+Quality+Control:+Statistical+Techniques,+edited+by+Edward+&ots=SFL0P5ZRVn&sig=dUG-VobzUwY1eZ6_SClpDGuNwKM&redir_esc=y#v=onepage&q&f=false.

International Energy Agency. 2021. "Portugal 2021." IEA. July 2021. https://www.iea.org/reports/portugal-2021.

International Energy Agency. 2022. “Portugal Electricity Security Policy.” IEA. June 30, 2022. https://www.iea.org/articles/portugal-electricity-security-policy.

Junfeng Qiao, Aihua Zhou, Lin Peng, Hua Gu, Xiaofeng Shen, Chenhong Huang, "Research and implementation of key technologies for electric distribution data analysis," 2023 Proc. SPIE 12703. https://spie.org/Publications/Proceedings/Paper/10.1117/12.2682795?SSO=1.

Kramar, Vadim, and Vasiliy Alchakov. “Time-Series Forecasting of Seasonal Data Using Machine Learning Methods.” *Algorithms* 16, no. 5 (May 10, 2023): 248. https://doi.org/10.3390/a16050248.

Nogueira, Filipe Tomé Brites Ferreira. 2014. “Produção de Energia Por via Fotovoltaica Para Autoconsumo Instantâneo Com Ligação à Rede.” Universidade da Beira Interior.

Ribeiro, Paulo, José F. G. Mendes, and Rui a. R. Ramos. 2012. *Advances in Environment, Computational Chemistry and Bioscience: Electric Mobility in Portugal – Municipal Plans for Its Promotion*.

S. M. Amin, "Smart grid security, privacy, and resilient architectures: Opportunities and challenges," 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 2012, pp. 1-2, https://ieeexplore.ieee.org/abstract/document/6345767.

Sierzchula, William, Sjoerd Bakker, Kees Maat, and Bert Van Wee. 2014. “The Influence of Financial Incentives and Other Socio-Economic Factors on Electric Vehicle Adoption.” *Energy Policy* 68 (May): 183–94. https://doi.org/10.1016/j.enpol.2014.01.043.

Suspiro, Ana. 2023. “Produção renovável abasteceu sozinha consumo nacional de eletricidade por 149 horas (cerca de seis dias).” *Observador*, November 6, 2023. https://observador.pt/2023/11/06/producao-renovavel-abasteceu-sozinha-consumo-nacional-de-eletricidade-por-149-horas-cerca-de-seis-dias/.

V. Santos, A. J. C. Pereira and C. M. Machado Ferreira, "Effective and simple measures to improve the Electric Power Distribution systems service continuity," 2014 49th International Universities Power Engineering Conference (UPEC), Cluj-Napoca, Romania,2014,pp.1-6, https://ieeexplore.ieee.org/abstract/document/6934767?casa_token=cfQZfE_hRewAAAAA:cjXQCBwCAbXNwweMKzVVG1sZHvcguYWN2oLLk4usVbAMFwX5b9cOfjMLqjKl9W8FM1Zs5i1i9fE.

Wang, Yusen. "Short-Term Power Load Forecasting Based on Machine Learning." DIVA, 2020. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1447321&dswid=-1718.

Wee, Sherilyn, Makena Coffman, and Sumner La Croix. 2018. "Do Electric Vehicle Incentives Matter? Evidence from the 50 U.S. States." *Research Policy* 47 (9): 1601–10. https://doi.org/10.1016/j.respol.2018.05.003.

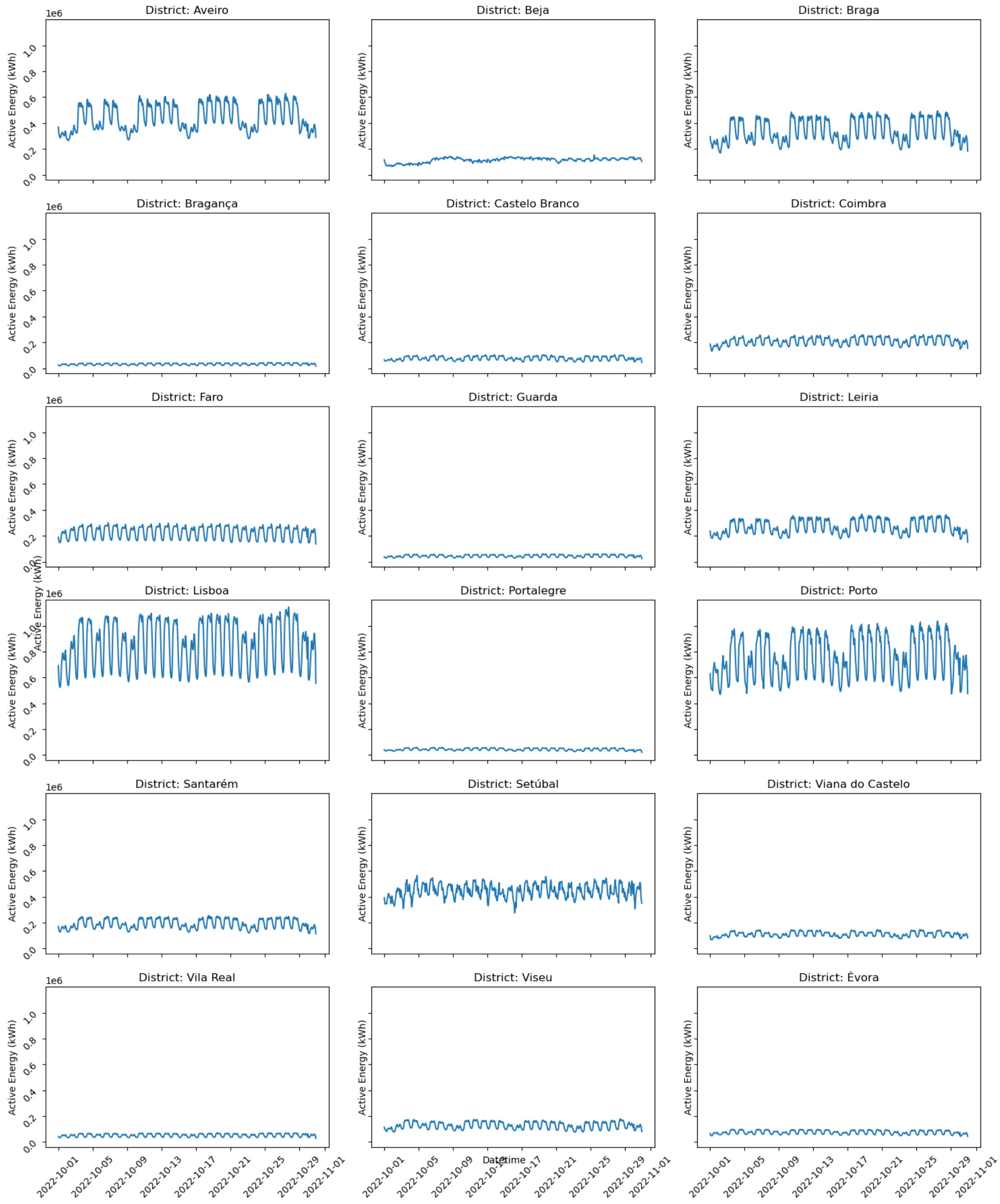
Zhang, Qi, Hailong Li, Lei Zhu, Pietro Elia Campana, Huihui Lu, Fredrik Wallin, and Qiang Sun. 2018. "Factors Influencing the Economics of Public Charging Infrastructures for EV – A Review." *Renewable & Sustainable Energy Review*

Zhou, Kaile, Chao Fu, and Shanlin Yang. "Big Data Driven Smart Energy Management: From Big Data to Big Insights." *Renewable & Sustainable Energy Reviews* 56 (April 1, 2016): 215–25. https://doi.org/10.1016/j.rser.2015.11.050.

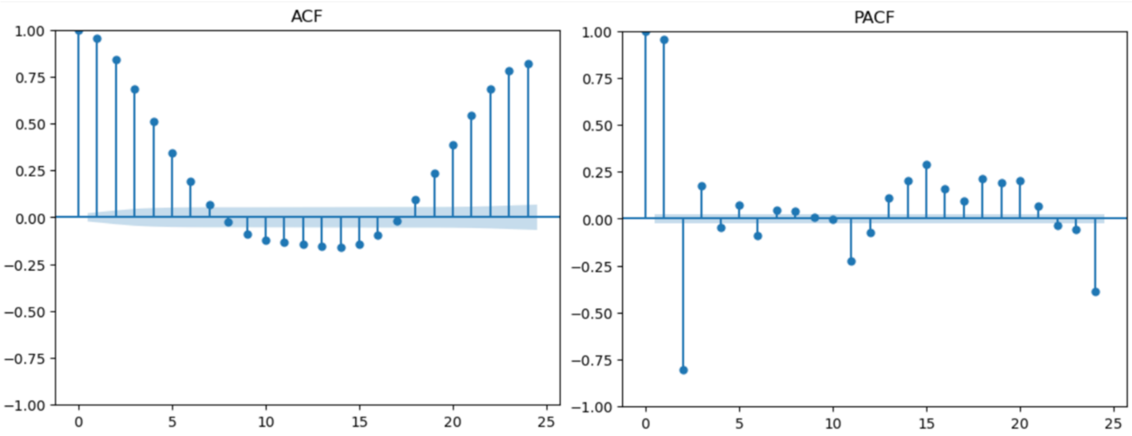
Zhou, Yan, Michael Wang, Han Hao, Linda Cooke Johnson, and Hewu Wang. 2014. "Plug-in Electric Vehicle Market Penetration and Incentives: A Global Review." *Mitigation and Adaptation Strategies for Global Change* 20 (5): 777–95. https://doi.org/10.1007/s11027-014-9611-2.

Appendix

Appendix A. Active Energy Time-Series Plot per District.



Appendix B. ACF and PACF plots



Appendix C. SARIMAX Models Summaries per District

```
SARIMA Model Summary for District Aveiro:
                                     SARIMAX Results
==================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:                 6432
Model:               SARIMAX(1, 1, [1], 24)   Log Likelihood              -69306.275
Date:                    Mon, 20 Nov 2023   AIC                         138634.550
Time:                            18:07:47   BIC                         138708.925
Sample:                        10-01-2022   HQIC                        138660.297
                             - 06-25-2023
Covariance Type:                      opg
=========================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
temperature_2m (°C)    55.0758     82.961      0.664      0.507    -107.526     217.677
precipitation (mm)    240.9725    373.767      0.645      0.519    -491.597     973.542
cloudcover (%)         18.7833      6.466      2.905      0.004       6.111      31.456
elevation               0.0005     25.526   1.79e-05      1.000     -50.029      50.030
lag_1                   1.4045      0.004    318.573      0.000       1.396       1.413
lag_2                  -0.4696      0.006    -77.702      0.000      -0.481      -0.458
day_type_weekday     1.051e+04    829.231     12.678      0.000    8887.917    1.21e+04
day_type_weekend    -1219.7402    818.128     -1.491      0.136   -2823.241     383.761
ar.S.L24                0.2462      0.010     25.798      0.000       0.228       0.265
ma.S.L24               -0.9071      0.006   -150.992      0.000      -0.919      -0.895
sigma2               1.903e+08      0.092   2.06e+09      0.000     1.9e+08     1.9e+08
===================================================================================
Ljung-Box (L1) (Q):                  26.79   Jarque-Bera (JB):             369354.51
Prob(Q):                              0.00   Prob(JB):                          0.00
Heteroskedasticity (H):               0.70   Skew:                              1.25
Prob(H) (two-sided):                  0.00   Kurtosis:                         40.18
===================================================================================
```

```
SARIMA Model Summary for District Beja:
                                     SARIMAX Results
==========================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:                   6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood                -60580.174
Date:                    Mon, 20 Nov 2023   AIC                           121182.347
Time:                            18:10:42   BIC                           121256.723
Sample:                        10-01-2022   HQIC                          121208.095
                             - 06-25-2023
Covariance Type:                      opg
==============================================================================================
                         coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
temperature_2m (°C)    13.5748     21.209      0.640      0.522     -27.995      55.145
precipitation (mm)    305.6684    313.742      0.974      0.330    -309.254     920.591
cloudcover (%)          0.3700      1.781      0.208      0.835      -3.121       3.861
elevation              41.3180      5.238      7.887      0.000      31.051      51.585
lag_1                   1.0320      0.002    437.505      0.000       1.027       1.037
lag_2                  -0.0957      0.004    -23.484      0.000      -0.104      -0.088
day_type_weekday     -366.1302    224.921     -1.628      0.104    -806.968      74.707
day_type_weekend      107.9937    236.310      0.457      0.648    -355.166     571.153
ar.S.L24                0.1141      0.008     15.198      0.000       0.099       0.129
ma.S.L24               -0.8929      0.004   -229.543      0.000      -0.901      -0.885
sigma2               1.179e+07      0.192   6.13e+07      0.000    1.18e+07    1.18e+07
===================================================================================
Ljung-Box (L1) (Q):                   1.95   Jarque-Bera (JB):            3425181.69
Prob(Q):                              0.16   Prob(JB):                          0.00
Heteroskedasticity (H):               0.86   Skew:                              0.55
Prob(H) (two-sided):                  0.00   Kurtosis:                        116.48
===================================================================================

SARIMA Model Summary for District Braga:
                                     SARIMAX Results
==========================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:                   6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood                -68445.966
Date:                    Mon, 20 Nov 2023   AIC                           136913.933
Time:                            18:11:47   BIC                           136988.308
Sample:                        10-01-2022   HQIC                          136939.680
                             - 06-25-2023
Covariance Type:                      opg
==============================================================================================
                         coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
temperature_2m (°C)    63.5817     70.582      0.901      0.368     -74.756     201.920
precipitation (mm)    128.3615    246.456      0.521      0.602    -354.683     611.406
cloudcover (%)         13.4243      5.483      2.448      0.014       2.678      24.170
elevation               0.0003     12.797   2.39e-05      1.000     -25.082      25.083
lag_1                   1.4289      0.005    270.844      0.000       1.419       1.439
lag_2                  -0.4976      0.006    -78.610      0.000      -0.510      -0.485
day_type_weekday     8625.4290    695.045     12.410      0.000    7263.165    9987.693
day_type_weekend    -1008.0674    684.936     -1.472      0.141   -2350.517     334.382
ar.S.L24                0.2599      0.011     23.865      0.000       0.239       0.281
ma.S.L24               -0.9118      0.006   -145.284      0.000      -0.924      -0.899
sigma2               1.471e+08      0.115   1.28e+09      0.000    1.47e+08    1.47e+08
===================================================================================
Ljung-Box (L1) (Q):                  16.12   Jarque-Bera (JB):             123685.71
Prob(Q):                              0.00   Prob(JB):                          0.00
Heteroskedasticity (H):               0.75   Skew:                              0.76
Prob(H) (two-sided):                  0.00   Kurtosis:                         24.51
===================================================================================
```

```
SARIMA Model Summary for District Bragança:
                               SARIMAX Results
==================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:               6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood            -53121.286
Date:                    Mon, 20 Nov 2023   AIC                       106264.571
Time:                            18:14:48   BIC                       106338.947
Sample:                        10-01-2022   HQIC                      106290.319
                             - 06-25-2023
Covariance Type:                      opg
=======================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------
temperature_2m (°C)   -20.2090      5.953     -3.395      0.001     -31.877      -8.541
precipitation (mm)     79.3518     43.875      1.809      0.071      -6.641     165.345
cloudcover (%)          0.6850      0.509      1.346      0.178      -0.312       1.682
elevation              12.1621      2.743      4.434      0.000       6.787      17.538
lag_1                   1.1627      0.004    284.696      0.000       1.155       1.171
lag_2                  -0.3075      0.006    -53.003      0.000      -0.319      -0.296
day_type_weekday      817.7973     62.554     13.073      0.000     695.194     940.401
day_type_weekend      131.7189     62.037      2.123      0.034      10.128     253.310
ar.S.L24                0.2011      0.009     23.134      0.000       0.184       0.218
ma.S.L24               -0.8870      0.007   -131.729      0.000      -0.900      -0.874
sigma2               1.155e+06   6889.804    167.589      0.000    1.14e+06    1.17e+06
===================================================================================
Ljung-Box (L1) (Q):                   6.43   Jarque-Bera (JB):           663335.24
Prob(Q):                              0.01   Prob(JB):                        0.00
Heteroskedasticity (H):               0.50   Skew:                           -0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                       52.94
===================================================================================
```

```
SARIMA Model Summary for District Castelo Branco:
                               SARIMAX Results
==================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:               6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood            -58635.476
Date:                    Mon, 20 Nov 2023   AIC                       117292.952
Time:                            18:16:30   BIC                       117367.327
Sample:                        10-01-2022   HQIC                      117318.700
                             - 06-25-2023
Covariance Type:                      opg
=======================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------
temperature_2m (°C)   -21.7786     14.357     -1.517      0.129     -49.918       6.361
precipitation (mm)     45.6097     95.144      0.479      0.632    -140.869     232.089
cloudcover (%)          1.4326      1.226      1.168      0.243      -0.971       3.836
elevation              17.4469      1.852      9.421      0.000      13.817      21.076
lag_1                   1.2459      0.005    269.658      0.000       1.237       1.255
lag_2                  -0.3521      0.007    -52.097      0.000      -0.365      -0.339
day_type_weekday     1756.3885    164.613     10.670      0.000    1433.754    2079.023
day_type_weekend      -86.3828    163.373     -0.529      0.597    -406.587     233.822
ar.S.L24                0.1722      0.010     17.169      0.000       0.153       0.192
ma.S.L24               -0.9182      0.006   -148.212      0.000      -0.930      -0.906
sigma2               6.702e+06      0.075   8.99e+07      0.000      6.7e+06     6.7e+06
===================================================================================
Ljung-Box (L1) (Q):                   7.55   Jarque-Bera (JB):           175866.00
Prob(Q):                              0.01   Prob(JB):                        0.00
Heteroskedasticity (H):               0.75   Skew:                            0.23
Prob(H) (two-sided):                  0.00   Kurtosis:                       28.71
===================================================================================
```

```
SARIMA Model Summary for District Coimbra:
                                      SARIMAX Results                                       
============================================================================================
Dep. Variable:            active_energy_kWh   No. Observations:                     6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood                  -62939.937
Date:                    Mon, 20 Nov 2023   AIC                             125901.873
Time:                            18:19:45   BIC                             125976.248
Sample:                        10-01-2022   HQIC                            125927.621
                             - 06-25-2023                                         
Covariance Type:                      opg                                         
=====================================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------
temperature_2m (°C)    -3.3612     29.906     -0.112      0.911     -61.976      55.253
precipitation (mm)    314.5249    142.598      2.206      0.027      35.037     594.012
cloudcover (%)          7.0446      2.543      2.771      0.006       2.061      12.028
elevation             117.9749      8.431     13.994      0.000     101.451     134.499
lag_1                   1.1655      0.003    371.589      0.000       1.159       1.172
lag_2                  -0.2639      0.006    -42.390      0.000      -0.276      -0.252
day_type_weekday     3720.9344    336.951     11.043      0.000    3060.523    4381.345
day_type_weekend      926.0602    339.936      2.724      0.006     259.797    1592.323
ar.S.L24                0.1511      0.008     18.133      0.000       0.135       0.167
ma.S.L24               -0.9136      0.006   -147.907      0.000      -0.926      -0.902
sigma2               2.528e+07      0.139   1.82e+08      0.000    2.53e+07    2.53e+07
===============================================================================================
Ljung-Box (L1) (Q):                   1.98   Jarque-Bera (JB):            3506622.75
Prob(Q):                              0.16   Prob(JB):                          0.00
Heteroskedasticity (H):               0.50   Skew:                              1.64
Prob(H) (two-sided):                  0.00   Kurtosis:                        117.78
===============================================================================================
```

```
SARIMA Model Summary for District Faro:
                                      SARIMAX Results                                       
============================================================================================
Dep. Variable:            active_energy_kWh   No. Observations:                     6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood                  -63150.346
Date:                    Mon, 20 Nov 2023   AIC                             126322.692
Time:                            18:20:40   BIC                             126397.068
Sample:                        10-01-2022   HQIC                            126348.440
                             - 06-25-2023                                         
Covariance Type:                      opg                                         
=====================================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------
temperature_2m (°C)   -21.6208     44.460     -0.486      0.627    -108.760      65.519
precipitation (mm)   1157.8477    458.517      2.525      0.012     259.172    2056.524
cloudcover (%)          4.0784      2.543      1.604      0.109      -0.906       9.063
elevation               0.0009     19.817   4.32e-05      1.000     -38.840      38.841
lag_1                   1.1155      0.002    501.946      0.000       1.111       1.120
lag_2                  -0.2016      0.005    -42.871      0.000      -0.211      -0.192
day_type_weekday     3214.7378    277.333     11.592      0.000    2671.176    3758.300
day_type_weekend      997.8720    283.920      3.515      0.000     441.399    1554.345
ar.S.L24                0.1712      0.007     23.241      0.000       0.157       0.186
ma.S.L24               -0.8091      0.006   -126.478      0.000      -0.822      -0.797
sigma2               2.471e+07      0.072   3.42e+08      0.000    2.47e+07    2.47e+07
===============================================================================================
Ljung-Box (L1) (Q):                  12.94   Jarque-Bera (JB):            3190172.00
Prob(Q):                              0.00   Prob(JB):                          0.00
Heteroskedasticity (H):               0.54   Skew:                              0.21
Prob(H) (two-sided):                  0.00   Kurtosis:                        112.52
===============================================================================================
```

```
SARIMA Model Summary for District Guarda:
                                    SARIMAX Results
==================================================================================
Dep. Variable:           active_energy_kWh   No. Observations:                 6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood              -54407.504
Date:                     Mon, 20 Nov 2023   AIC                         108837.008
Time:                             18:23:35   BIC                         108911.383
Sample:                         10-01-2022   HQIC                        108862.756
                              - 06-25-2023
Covariance Type:                       opg
=========================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
temperature_2m (°C)    -8.1364      7.693     -1.058      0.290     -23.214       6.941
precipitation (mm)     94.1919     56.668      1.662      0.096     -16.875     205.259
cloudcover (%)          0.6838      0.668      1.024      0.306      -0.625       1.993
elevation               7.8349      1.694      4.625      0.000       4.514      11.155
lag_1                   1.2701      0.004    288.566      0.000       1.261       1.279
lag_2                  -0.3858      0.005    -73.297      0.000      -0.396      -0.375
day_type_weekday     1099.6519     82.914     13.263      0.000     937.143    1262.161
day_type_weekend        8.0082     79.996      0.100      0.920    -148.782     164.798
ar.S.L24                0.1922      0.008     23.340      0.000       0.176       0.208
ma.S.L24               -0.9068      0.007   -139.432      0.000      -0.920      -0.894
sigma2               1.786e+06   1.13e+04    157.418      0.000    1.76e+06    1.81e+06
===================================================================================
Ljung-Box (L1) (Q):                  14.09   Jarque-Bera (JB):            845025.04
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.51   Skew:                            -0.36
Prob(H) (two-sided):                  0.00   Kurtosis:                        59.36
===================================================================================
```

```
SARIMA Model Summary for District Leiria:
                                    SARIMAX Results
==================================================================================
Dep. Variable:           active_energy_kWh   No. Observations:                 6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood              -65597.918
Date:                     Mon, 20 Nov 2023   AIC                         131217.836
Time:                             18:26:04   BIC                         131292.211
Sample:                         10-01-2022   HQIC                        131243.584
                              - 06-25-2023
Covariance Type:                       opg
=========================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
temperature_2m (°C)    37.7015     56.361      0.669      0.504     -72.763     148.166
precipitation (mm)    141.2568    357.677      0.395      0.693    -559.777     842.291
cloudcover (%)         19.5770      3.810      5.138      0.000      12.109      27.045
elevation             115.4491     11.143     10.361      0.000      93.610     137.288
lag_1                   1.3246      0.004    346.259      0.000       1.317       1.332
lag_2                  -0.4134      0.006    -70.320      0.000      -0.425      -0.402
day_type_weekday     6000.3394    471.322     12.731      0.000    5076.565    6924.114
day_type_weekend     -357.8138    469.233     -0.763      0.446   -1277.494     561.867
ar.S.L24                0.2381      0.008     28.895      0.000       0.222       0.254
ma.S.L24               -0.9234      0.005   -180.334      0.000      -0.933      -0.913
sigma2               6.027e+07      0.086   7.01e+08      0.000    6.03e+07    6.03e+07
===================================================================================
Ljung-Box (L1) (Q):                  19.67   Jarque-Bera (JB):            848423.52
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.62   Skew:                             0.97
Prob(H) (two-sided):                  0.00   Kurtosis:                        59.45
===================================================================================
```

```
SARIMA Model Summary for District Lisboa:
                                    SARIMAX Results
==========================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:                   6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood                -71824.964
Date:                    Mon, 20 Nov 2023   AIC                           143671.928
Time:                            18:27:01   BIC                           143746.304
Sample:                        10-01-2022   HQIC                          143697.676
                             - 06-25-2023
Covariance Type:                      opg
=================================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
temperature_2m (°C)   316.6685    180.193      1.757      0.079     -36.503     669.840
precipitation (mm)     34.4558   1119.043      0.031      0.975   -2158.828    2227.740
cloudcover (%)         20.8917     10.028      2.083      0.037       1.238      40.545
elevation            7.876e-05     68.533   1.15e-06      1.000    -134.321     134.321
lag_1                   1.3575      0.002    605.040      0.000       1.353       1.362
lag_2                  -0.4465      0.004    -99.557      0.000      -0.455      -0.438
day_type_weekday     1.565e+04   1145.704     13.663      0.000    1.34e+04    1.79e+04
day_type_weekend      915.7929   1171.123      0.782      0.434   -1379.566    3211.151
ar.S.L24                0.2209      0.007     33.416      0.000       0.208       0.234
ma.S.L24               -0.8290      0.005   -158.454      0.000      -0.839      -0.819
sigma2               3.882e+08      0.039   1.01e+10      0.000    3.88e+08    3.88e+08
==========================================================================================
Ljung-Box (L1) (Q):                  23.81   Jarque-Bera (JB):             4935841.34
Prob(Q):                              0.00   Prob(JB):                           0.00
Heteroskedasticity (H):               0.38   Skew:                               2.10
Prob(H) (two-sided):                  0.00   Kurtosis:                         139.17
==========================================================================================
```

```
SARIMA Model Summary for District Portalegre:
                                    SARIMAX Results
==========================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:                   6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood                -53830.005
Date:                    Mon, 20 Nov 2023   AIC                           107682.010
Time:                            18:59:54   BIC                           107756.385
Sample:                        10-01-2022   HQIC                          107707.757
                             - 06-25-2023
Covariance Type:                      opg
=================================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
temperature_2m (°C)    -1.6003      7.238     -0.221      0.825     -15.787      12.587
precipitation (mm)     35.6241     51.294      0.695      0.487     -64.910     136.158
cloudcover (%)          0.8769      0.592      1.482      0.138      -0.283       2.037
elevation              17.8315      3.635      4.905      0.000      10.706      24.957
lag_1                   1.2226      0.004    304.241      0.000       1.215       1.230
lag_2                  -0.3204      0.005    -58.516      0.000      -0.331      -0.310
day_type_weekday      883.4391     76.531     11.544      0.000     733.441    1033.438
day_type_weekend      -67.9285     74.806     -0.908      0.364    -214.546      78.689
ar.S.L24                0.2013      0.007     27.101      0.000       0.187       0.216
ma.S.L24               -0.9057      0.006   -150.510      0.000      -0.917      -0.894
sigma2                1.43e+06   7866.839    181.767      0.000    1.41e+06    1.45e+06
==========================================================================================
Ljung-Box (L1) (Q):                  13.15   Jarque-Bera (JB):              857343.51
Prob(Q):                              0.00   Prob(JB):                           0.00
Heteroskedasticity (H):               0.67   Skew:                               0.20
Prob(H) (two-sided):                  0.00   Kurtosis:                          59.78
==========================================================================================
```

```
SARIMA Model Summary for District Porto:
                                   SARIMAX Results
==================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:                 6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood              -73902.101
Date:                    Mon, 20 Nov 2023   AIC                         147826.202
Time:                            19:01:07   BIC                         147900.577
Sample:                        10-01-2022   HQIC                        147851.950
                             - 06-25-2023
Covariance Type:                      opg
=======================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------
temperature_2m (°C)   -17.4557    168.330     -0.104      0.917    -347.377     312.465
precipitation (mm)    514.7873    686.550      0.750      0.453    -830.826    1860.400
cloudcover (%)         40.3143     12.450      3.238      0.001      15.912      64.717
elevation               0.0004     36.914   9.99e-06      1.000     -72.350      72.351
lag_1                   1.1586      0.005    210.968      0.000       1.148       1.169
lag_2                  -0.2586      0.008    -31.854      0.000      -0.274      -0.243
day_type_weekday      2.04e+04   1650.525     12.360      0.000    1.72e+04    2.36e+04
day_type_weekend      941.6687   1568.970      0.600      0.548   -2133.456    4016.794
ar.S.L24                0.2104      0.011     19.374      0.000       0.189       0.232
ma.S.L24               -0.8706      0.007   -116.273      0.000      -0.885      -0.856
sigma2               7.607e+08      0.069   1.11e+10      0.000    7.61e+08    7.61e+08
===================================================================================
Ljung-Box (L1) (Q):                  10.03   Jarque-Bera (JB):              64137.71
Prob(Q):                              0.00   Prob(JB):                          0.00
Heteroskedasticity (H):               0.66   Skew:                              0.43
Prob(H) (two-sided):                  0.00   Kurtosis:                         18.51
===================================================================================
```

```
SARIMA Model Summary for District Santarém:
                                   SARIMAX Results
==================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:                 6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood              -63286.026
Date:                    Mon, 20 Nov 2023   AIC                         126594.053
Time:                            19:02:10   BIC                         126668.428
Sample:                        10-01-2022   HQIC                        126619.801
                             - 06-25-2023
Covariance Type:                      opg
=======================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------
temperature_2m (°C)    41.0866     32.948      1.247      0.212     -23.491     105.664
precipitation (mm)    131.8955    229.822      0.574      0.566    -318.548     582.339
cloudcover (%)          5.3820      2.385      2.257      0.024       0.708      10.056
elevation               0.0004     13.882   3.18e-05      1.000     -27.207      27.208
lag_1                   1.2919      0.004    341.459      0.000       1.285       1.299
lag_2                  -0.3758      0.006    -64.955      0.000      -0.387      -0.364
day_type_weekday     4793.7362    320.477     14.958      0.000    4165.614    5421.859
day_type_weekend       28.3225    308.291      0.092      0.927    -575.918     632.563
ar.S.L24                0.2433      0.009     26.097      0.000       0.225       0.262
ma.S.L24               -0.8883      0.007   -128.052      0.000      -0.902      -0.875
sigma2               2.813e+07      0.100   2.82e+08      0.000    2.81e+07    2.81e+07
===================================================================================
Ljung-Box (L1) (Q):                  24.30   Jarque-Bera (JB):             538659.98
Prob(Q):                              0.00   Prob(JB):                          0.00
Heteroskedasticity (H):               0.58   Skew:                              0.90
Prob(H) (two-sided):                  0.00   Kurtosis:                         47.97
===================================================================================
```

```
SARIMA Model Summary for District Setúbal:
                                      SARIMAX Results
==========================================================================================
Dep. Variable:            active_energy_kWh   No. Observations:                   6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood                  -72962.256
Date:                      Mon, 20 Nov 2023   AIC                             145946.513
Time:                              19:03:39   BIC                             146020.888
Sample:                          10-01-2022   HQIC                            145972.260
                               - 06-25-2023
Covariance Type:                        opg
=========================================================================================
                         coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
temperature_2m (°C)   137.2620    148.213      0.926      0.354    -153.230     427.754
precipitation (mm)   1422.6799   1360.963      1.045      0.296   -1244.759    4090.119
cloudcover (%)         13.6806     10.974      1.247      0.213      -7.828      35.189
elevation               0.0003    103.483   2.46e-06      1.000    -202.824     202.824
lag_1                   0.9692      0.007    131.939      0.000       0.955       0.984
lag_2                  -0.0456      0.009     -4.884      0.000      -0.064      -0.027
day_type_weekday     6453.4615   1705.805      3.783      0.000    3110.145    9796.778
day_type_weekend     2855.0334   1780.984      1.603      0.109    -635.630    6345.697
ar.S.L24                0.1476      0.012     12.401      0.000       0.124       0.171
ma.S.L24               -0.9248      0.006   -154.062      0.000      -0.937      -0.913
sigma2               5.452e+08      0.015   3.66e+10      0.000    5.45e+08    5.45e+08
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              14068.95
Prob(Q):                              0.97   Prob(JB):                          0.00
Heteroskedasticity (H):               0.71   Skew:                             -0.07
Prob(H) (two-sided):                  0.00   Kurtosis:                         10.27
===================================================================================
```

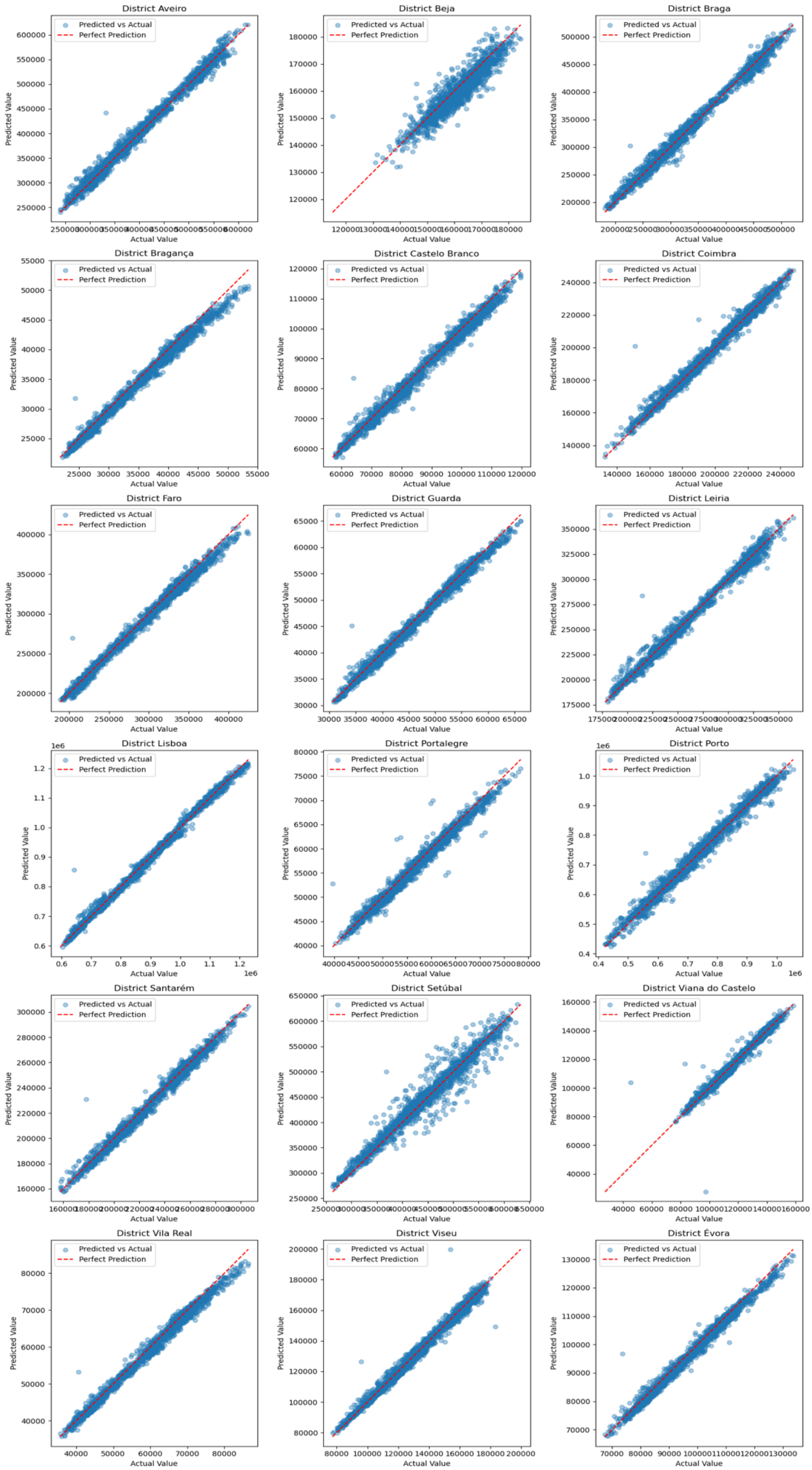
```
SARIMA Model Summary for District Viana do Castelo:
                                      SARIMAX Results
==========================================================================================
Dep. Variable:            active_energy_kWh   No. Observations:                   6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood                  -60337.616
Date:                      Mon, 20 Nov 2023   AIC                             120697.233
Time:                              19:04:38   BIC                             120771.608
Sample:                          10-01-2022   HQIC                            120722.980
                               - 06-25-2023
Covariance Type:                        opg
=========================================================================================
                         coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
temperature_2m (°C)    38.1937     21.105      1.810      0.070      -3.171      79.558
precipitation (mm)     -2.3582     64.189     -0.037      0.971    -128.167     123.450
cloudcover (%)          2.0100      1.577      1.275      0.202      -1.080       5.101
elevation               0.0007     10.557   7.05e-05      1.000     -20.691      20.692
lag_1                   1.2544      0.003    385.499      0.000       1.248       1.261
lag_2                  -0.3232      0.006    -55.058      0.000      -0.335      -0.312
day_type_weekday     1871.4149    212.896      8.790      0.000    1454.146    2288.684
day_type_weekend     -118.9134    210.495     -0.565      0.572    -531.475     293.648
ar.S.L24                0.1385      0.009     15.947      0.000       0.121       0.156
ma.S.L24               -0.9069      0.005   -173.531      0.000      -0.917      -0.897
sigma2               1.105e+07      0.059   1.88e+08      0.000    1.11e+07    1.11e+07
===================================================================================
Ljung-Box (L1) (Q):                   5.84   Jarque-Bera (JB):            1376439.55
Prob(Q):                              0.02   Prob(JB):                          0.00
Heteroskedasticity (H):               0.46   Skew:                              1.67
Prob(H) (two-sided):                  0.00   Kurtosis:                         74.86
===================================================================================
```

```
SARIMA Model Summary for District Vila Real:
                                     SARIMAX Results                                      
==========================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:                  6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood              -55643.631
Date:                    Mon, 20 Nov 2023   AIC                         111309.263
Time:                            19:07:34   BIC                         111383.638
Sample:                        10-01-2022   HQIC                        111335.011
                             - 06-25-2023                                         
Covariance Type:                      opg                                         
=====================================================================================
                        coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
temperature_2m (°C)   -24.8046      9.599     -2.584      0.010     -43.618      -5.991
precipitation (mm)    21.2637     40.402      0.526      0.599     -57.923     100.450
cloudcover (%)         0.6008      0.772      0.778      0.437      -0.913       2.115
elevation             13.7809      2.083      6.616      0.000       9.698      17.864
lag_1                  1.2299      0.004    286.720      0.000       1.221       1.238
lag_2                 -0.3572      0.006    -62.592      0.000      -0.368      -0.346
day_type_weekday    1174.5595     97.256     12.077      0.000     983.941    1365.178
day_type_weekend     205.5922     96.108      2.139      0.032      17.224     393.961
ar.S.L24               0.2154      0.009     23.955      0.000       0.198       0.233
ma.S.L24              -0.8986      0.007   -135.872      0.000      -0.912      -0.886
sigma2              2.611e+06   1.65e+04    157.780      0.000    2.58e+06    2.64e+06
===================================================================================
Ljung-Box (L1) (Q):                  10.80   Jarque-Bera (JB):            710235.40
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.47   Skew:                            -0.31
Prob(H) (two-sided):                  0.00   Kurtosis:                        54.67
===================================================================================
```

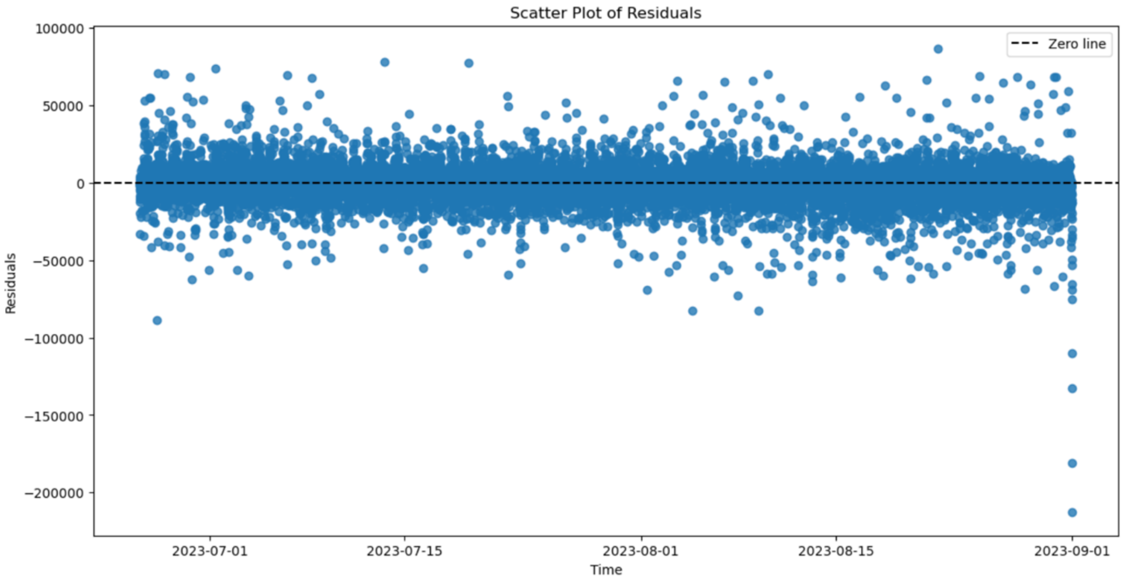
```
SARIMA Model Summary for District Viseu:
                                     SARIMAX Results                                      
==========================================================================================
Dep. Variable:          active_energy_kWh   No. Observations:                  6432
Model:             SARIMAX(1, 1, [1], 24)   Log Likelihood              -61491.698
Date:                    Mon, 20 Nov 2023   AIC                         123005.396
Time:                            19:09:50   BIC                         123079.771
Sample:                        10-01-2022   HQIC                        123031.144
                             - 06-25-2023                                         
Covariance Type:                      opg                                         
=====================================================================================
                        coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
temperature_2m (°C)     6.0267     23.245      0.259      0.795     -39.532      51.585
precipitation (mm)   117.8297    123.590      0.953      0.340    -124.402     360.062
cloudcover (%)         1.6745      1.923      0.871      0.384      -2.094       5.443
elevation             19.3631      2.289      8.458      0.000      14.876      23.850
lag_1                  1.3039      0.004    330.886      0.000       1.296       1.312
lag_2                 -0.3866      0.006    -62.604      0.000      -0.399      -0.374
day_type_weekday    2490.6242    232.734     10.702      0.000    2034.473    2946.775
day_type_weekend     172.3760    232.294      0.742      0.458    -282.911     627.663
ar.S.L24               0.2337      0.009     26.944      0.000       0.217       0.251
ma.S.L24              -0.9213      0.006   -159.539      0.000      -0.933      -0.910
sigma2               1.65e+07      0.109   1.51e+08      0.000    1.65e+07    1.65e+07
===================================================================================
Ljung-Box (L1) (Q):                  10.39   Jarque-Bera (JB):            546045.58
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.67   Skew:                             0.73
Prob(H) (two-sided):                  0.00   Kurtosis:                        48.29
===================================================================================
```

```
SARIMA Model Summary for District Évora:
                                      SARIMAX Results
==================================================================================
Dep. Variable:            active_energy_kWh   No. Observations:               6432
Model:               SARIMAX(1, 1, [1], 24)   Log Likelihood            -57099.140
Date:                      Mon, 20 Nov 2023   AIC                       114220.279
Time:                              19:13:04   BIC                       114294.655
Sample:                          10-01-2022   HQIC                      114246.027
                               - 06-25-2023
Covariance Type:                        opg
=======================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------
temperature_2m (°C)    17.0698     12.716      1.342      0.179      -7.853      41.992
precipitation (mm)     46.9387    136.349      0.344      0.731    -220.301     314.179
cloudcover (%)          0.8558      0.977      0.876      0.381      -1.060       2.771
elevation              22.2974      3.163      7.050      0.000      16.098      28.496
lag_1                   1.2817      0.005    284.629      0.000       1.273       1.290
lag_2                  -0.3799      0.006    -61.827      0.000      -0.392      -0.368
day_type_weekday     1425.4062    128.910     11.057      0.000    1172.748    1678.065
day_type_weekend        9.9528    124.301      0.080      0.936    -233.673     253.579
ar.S.L24                0.2512      0.009     28.501      0.000       0.234       0.268
ma.S.L24               -0.9163      0.007   -139.171      0.000      -0.929      -0.903
sigma2               4.145e+06   2.51e+04    164.947      0.000     4.1e+06    4.19e+06
===================================================================================
Ljung-Box (L1) (Q):                  22.87   Jarque-Bera (JB):           1192928.63
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.68   Skew:                             1.08
Prob(H) (two-sided):                  0.00   Kurtosis:                        69.94
===================================================================================
```

Appendix D. Comparison of Actual vs. Predicted Active Energy for Each District by SARIMAX Model.



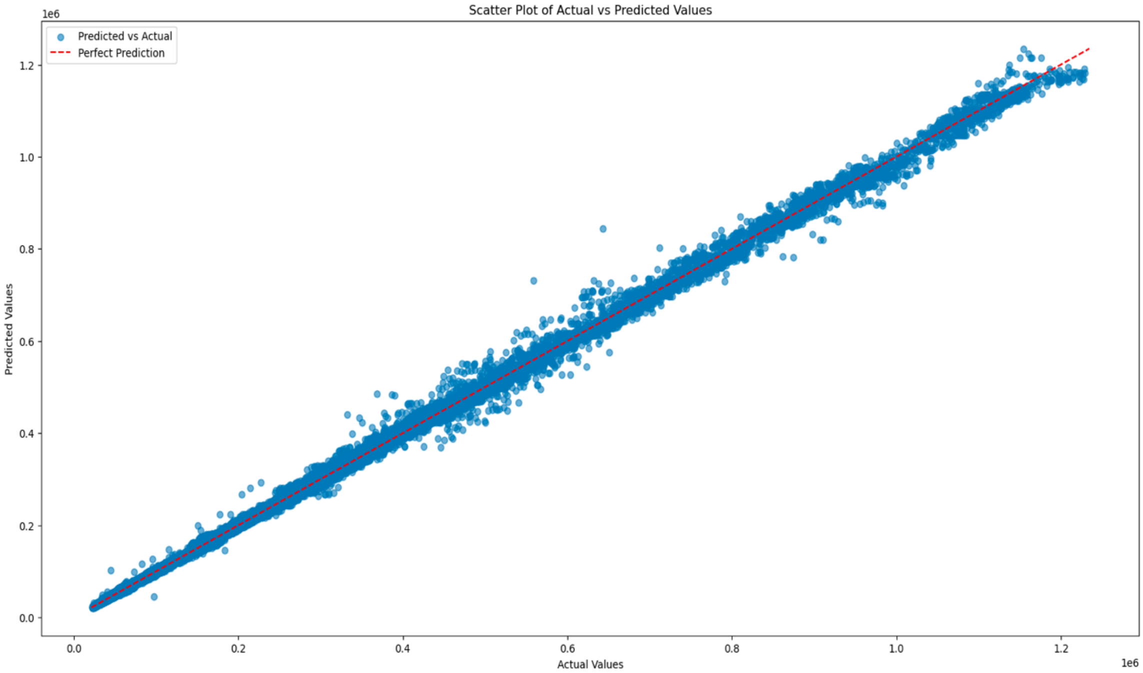
Appendix E. Scatter Plot of SARIMAX Residuals.



Appendix F. XGBoost Model Feature's Weight Importance.

	Feature	Score		Feature	Score
0	lag_1	434.0	**12**	district_Porto	11.0
1	hour	379.0	**13**	district_Lisboa	10.0
2	lag_2	183.0	**14**	district_Beja	9.0
3	temperature_2m (°C)	96.0	**15**	year	9.0
4	month	53.0	**16**	district_Leiria	7.0
5	day	53.0	**17**	day_type_weekend	6.0
6	elevation	43.0	**18**	district_Faro	4.0
7	day_type_weekday	31.0	**19**	district_Viana do Castelo	4.0
8	cloudcover (%)	31.0	**20**	district_Viseu	2.0
9	precipitation (mm)	17.0	**21**	district_Bragança	1.0
10	district_Aveiro	13.0	**22**	district_Castelo Branco	1.0
11	district_Braga	11.0			

Appendix G. Comparison of Actual vs. Predicted Active Energy by XGBoost Model.



Appendix H. Scatter Plot of XGBoost Residuals.