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 – RENEWABLE ENERGY INTEGRATION NETWORK

MARIANA MARÇALO ALVES

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

## Abbreviations

| | |
|---|---|
| **ACEA** | European Association of Automobile Manufacturers |
| **AI** | Artificial Intelligence |
| **APREN** | Portuguese Renewable Energy Association |
| **CPE** | Charging Point Equipment |
| **DGEG** | General Direction of Energy and Geology |
| **DSM** | Demand-Side Management |
| **DSO** | Distribution System Operator |
| **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 |
| **kV** | Kilovolt |
| **kW** | Kilowatt |
| **MAE** | Mean Absolute Error |
| **ML** | Machine Learning |
| **MSE** | Mean Squared Error |
| **MW** | Megawatts |
| **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 |
| **SARIMAX** | Seasonal Autoregressive Integrated Moving Average with Exogenous Factors |
| **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).

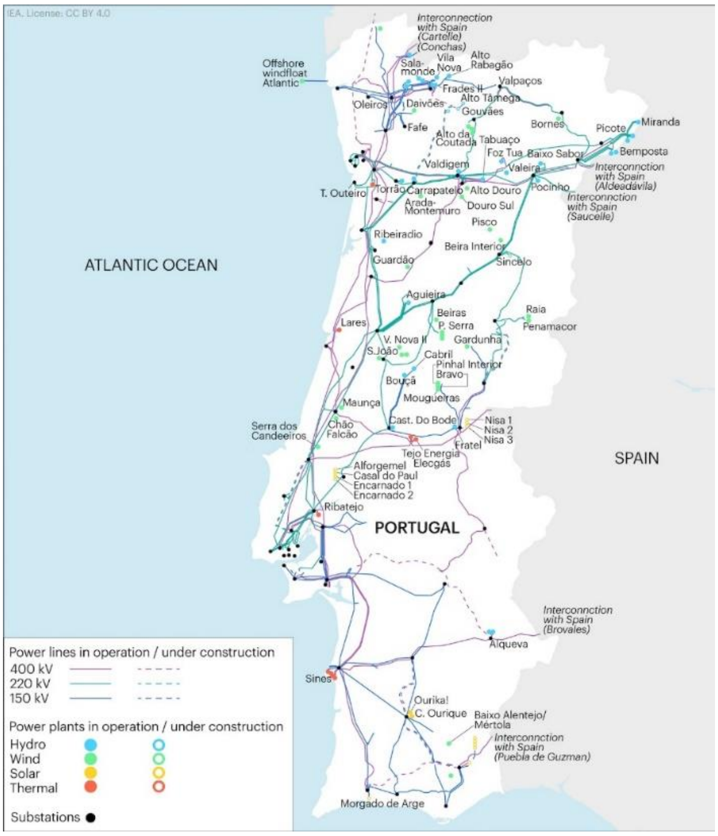


Figure 1. Portugal National Transport Grid Map. Source: REN.

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).

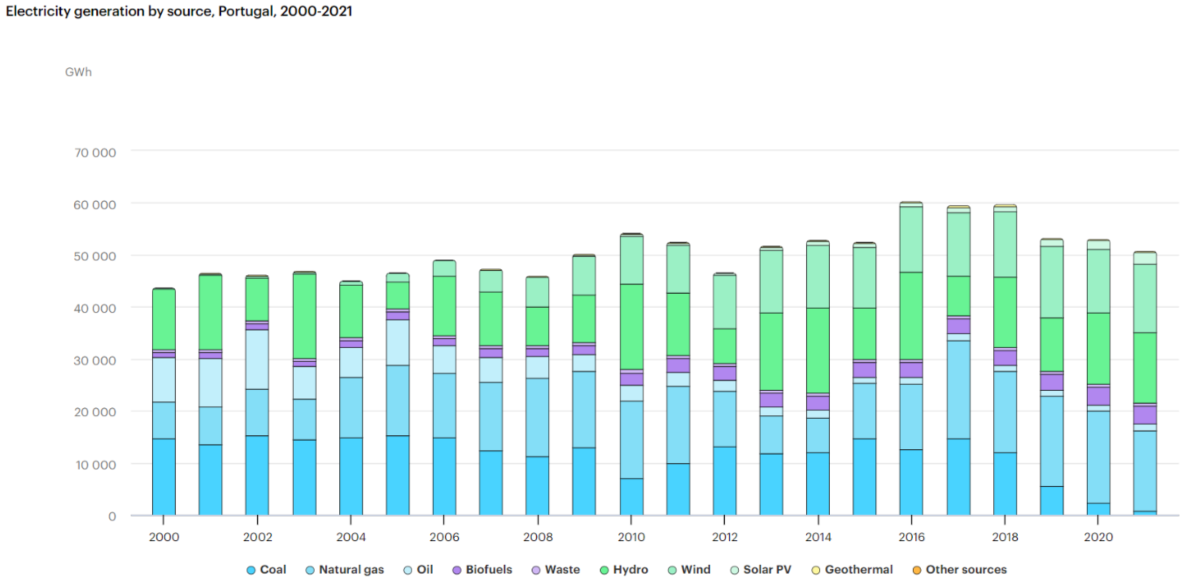


Figure 2. Electricity Generation by Source, Portugal, 2000-2021. Source: IEA.

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).

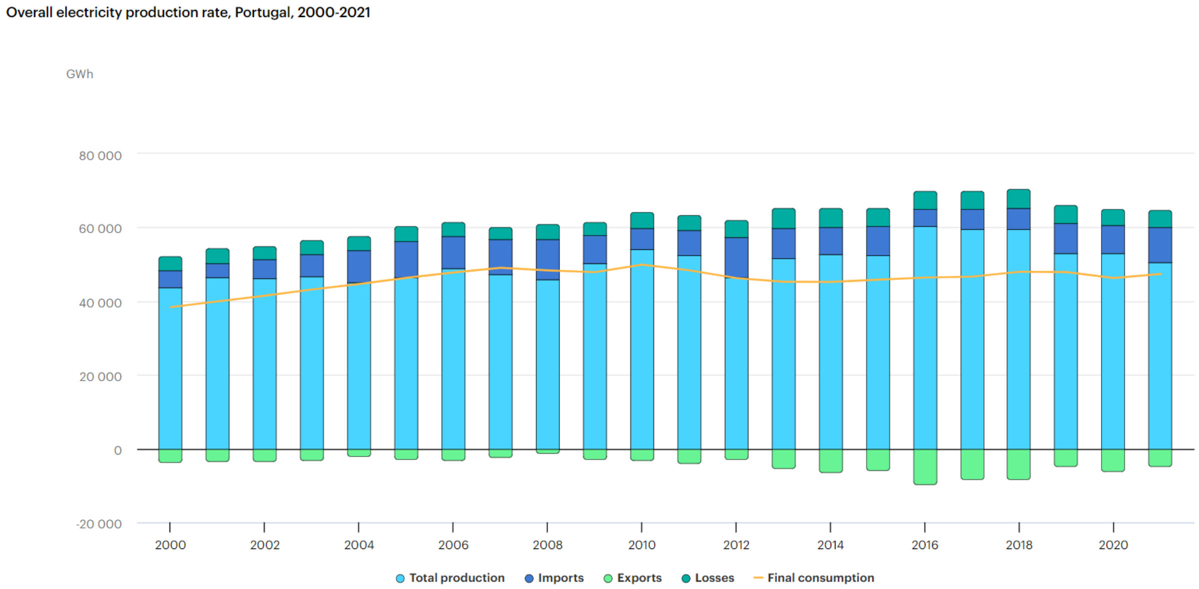


Figure 3. Overall Electricity Production Rate, Portugal, 2000-2021. Source: IEA.

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.

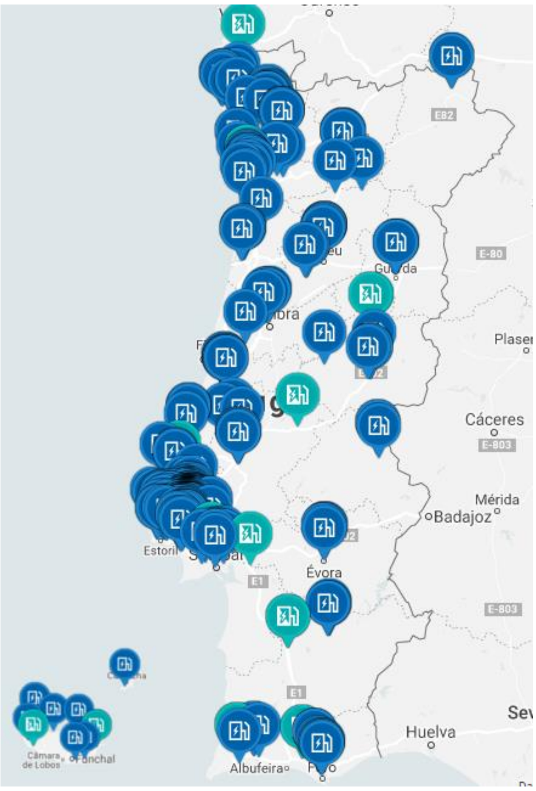


Figure 4. Distribution of CPEs in Portugal.

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.

## 6. Renewable Energy Integration Network

### 6.1. Introduction

Renewable energy adoption stands at the forefront of global initiatives to mitigate climate change and transition towards sustainable energy landscapes. In this context, exploring UPACs in Portugal is an important topic. UPACs, as decentralized renewable energy systems, embody the potential to reshape Portugal's energy landscape. This study aims to contribute with valuable insights to this ongoing discourse, illuminating the influences that govern the adoption of UPACs in Portugal's municipalities.

Focusing on the environmental, meteorological, and consumption factors, this research delves into the landscape of UPAC adoption in Portugal. The research question guiding this exploration is: What factors influence the growth and adoption of UPACs in Portugal's municipalities? The motivation stems from the need to unravel the unique regional characteristics influencing UPAC adoption and provide practical insights for policymakers, informing sustainable energy policies.

The research uncovers positive coefficients for meteorological variables in the models, emphasizing their significant influence on UPAC adoption rates. Shortwave radiation, direct radiation, and direct normal irradiance play fundamental roles, highlighting the importance of sunlight exposure in driving UPAC installations. Additionally, non-linear models such as XGBoost and Random Forest outperform linear models, underlining the complex, non-linear dynamics of the relationship between predictors and UPAC adoption.

The unfolding chapters of this thesis follow a structured path, starting with a meticulous exploration of the dataset's origins and the rationale behind the variable selection. The subsequent sections introduce a diverse set of ML models strategically chosen to capture the complexity of UPAC adoption factors. Rigorous model evaluation, hypothesis testing, and an in-depth examination of coefficients are the analytical backbone, providing a comprehensive understanding of the factors influencing UPAC adoption.

### 6.2. Literature Review

Self-consumption of energy has become a topic of interest in Portugal, significantly when electricity and gas prices have increased. Portugal is committed to achieving carbon neutrality by 2050 as a contribution to the European goals assumed in implementing the Paris Agreement. To support these objectives, the Environmental Fund’s More Sustainable Buildings program for 2022 allocated a budget of 45 million euros, which was later increased to 60 million euros, to incentivize investments in equipment capable of making properties more sustainable in energy, such as solar panels. The government aims to reduce energy consumption in intervened buildings by up to 30%.

Portugal advanced its target for incorporating renewable energy into electricity production by four years. From 2026 onward, 80% of the nation’s energy production will be from renewable sources, a goal initially set for 2030. With this, Portugal is preparing to achieve climate neutrality in 2045. The determination to be at the forefront of the energy transition sets ambitious goals for 2030, which were defined within the scope of the National Energy and Climate Plan for 2021-2030, including a 47% share of energy from renewable sources in final consumption gross and the requirement that at least 80% of electricity be produced from renewable sources. The plan emphasizes promoting local and decentralized clean energy production to reduce dependency on external energy sources.

To reinforce Portugal’s goal The Support Program for More Sustainable Buildings has been relaunched. The program, now with a budget of 30 million euros, aims to contribute to a minimum 30% reduction in primary energy consumption in the intervened buildings.

The demand for this type of solution has more than double in 2022. According to the data from the General Directorate of Energy and Geology, until July 2022, the installed power of production units for self-consumption reached 695 megawatts (MW), an increase of 103% compared to the same period of the previous year. However, the strong growth is due to the demand from individuals and the business sector. The president of the Portuguese Renewable Energy Association (APREN) believes that one of the factors for this increase is the support provided by the Environmental Fund for the acquisition of PV panels, which provided for reimbursement of up to 85%. This support aims to accelerate the energy transition by raising awareness among citizens about their active role in the decentralization of energy.

Integrating the Public Service Network (RESP) involves challenges and considerations for effectively incorporating renewable energy sources. Smart grid technology enhances grid flexibility, energy efficiency, and overall reliability. However, integrating variable energy sources poses challenges to the traditional electricity grid, initially designed for one-way power flow – from plants to consumers.

The challenges in integrating renewable energy into the power grid include intermittent and variable, impacting grid stability, infrastructure, and regulatory frameworks. For example, renewable energy sources like wind and solar are inherently intermittent and variable, leading to fluctuations in output based on weather conditions. There are also economic challenges, such as the need for new pricing models and incentives, that need the involvement of policymakers and regulators to create a favorable environment. Successful integrations require addressing these challenges to ensure grid stability, reliability, and efficiency in transitioning to a more sustainable energy landscape.

The growing demand for self-consumption solutions in Portugal is reflected in the presence of over 131 thousand production units for self-consumption of energy across the country. Guimarães stands out as the leader, with 3 650 installations. Seixal and Vila Nova de Famalicão follow closely, occupying second and third place in the ranking, with 3 192 and 3 190 installations of UPACs.

While not all UPACs exclusively harness solar energy, some incorporate micro-wind or biomass production, the national data from the DGEG reveals that PV panels dominate, constituting 98.7% of the installed power of UPACs. Interestingly, despite the southern region of Portugal having more significant solar energy potential, the northern part of the country exhibits a higher adoption of UPACs (Figure 5).

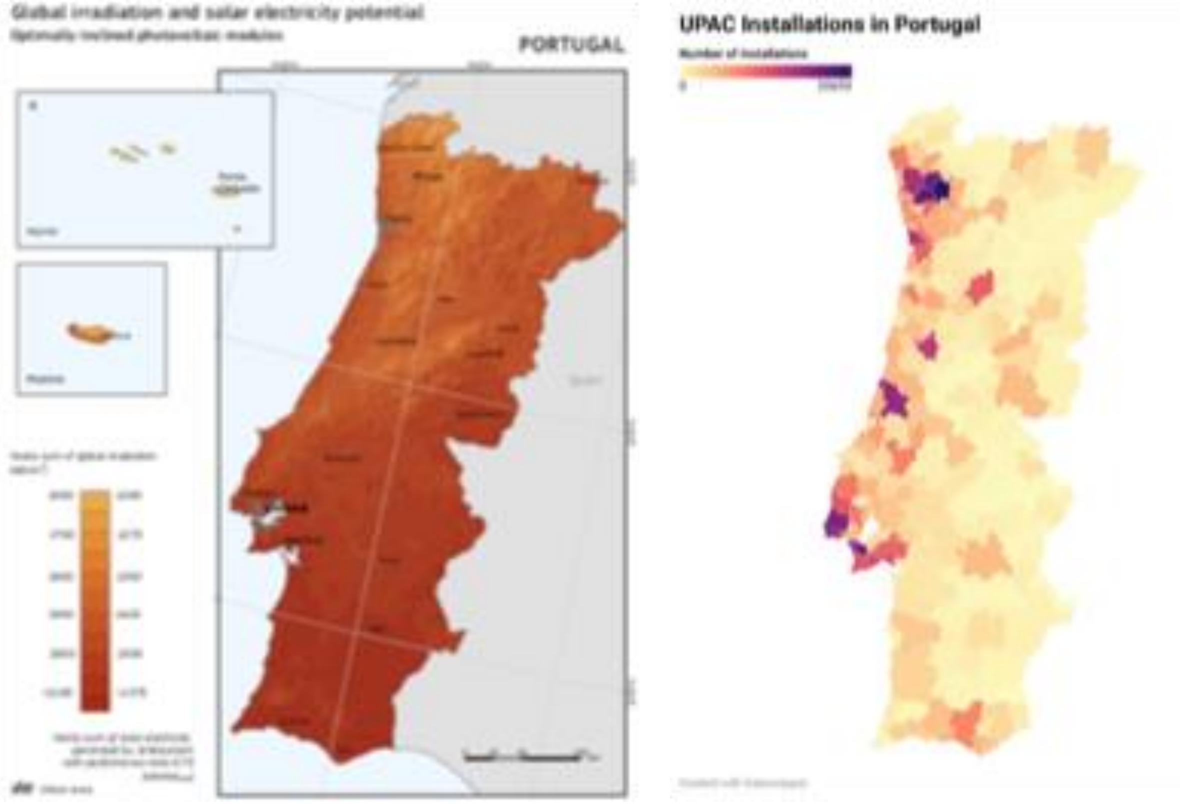


Figure 5. Global Irradiation and Solar Electricity Potential vs. UPAC Installations in Portugal. Source: European Commission, E-Redes.

Financial support has been extended to various entities to bolster the adoption of renewable energy systems. The municipality of Seixal has been financially supporting clubs and institutions to facilitate the installation of PV solar systems due to being a municipality with abundant annual sun exposure of over 3 thousand hours.

### 6.3. Research Questions and Hypothesis

This thesis investigates the predictive factors influencing the adoption of UPACs, particularly exploring their seamless integration into the Public Service Electric Network (RESP) across various municipalities. As such, this study's primary question is: What predictive factors can influence the growth and adoption of UPACs and their integration into the RESP in different municipalities?

To address the previous research question, the study formulates a hypothesis to delve into the specific factors that potentially impact the adoption of UPACs in municipalities. This first hypothesis focuses on the influence of environmental and meteorological factors on the adoption rates of UPACs. The hypothesis articulates as follows: The adoption of UPAC units in municipalities is influenced by environmental and meteorological factors (Shortwave Radiation, Direct Radiation, Diffuse Radiation, Direct Normal and Irradiance, Maximum Temperature, Minimum Temperature, Precipitation).

This hypothesis seeks to elucidate the role of environmental and weather-related variables in driving the uptake of UPACs, contributing to a more nuanced understanding of the dynamics at play in integrating renewable energy sources into the electric network infrastructure.

Furthermore, a second hypothesis proposes to evaluate whether supervised ML models have statistical significance in predicting the number of UPAC installations per municipality. The hypothesis is formulated as follows: All ML models are equally efficient in predicting the number of UPAC installations per municipality.

### 6.4. Methods

This study is essential for understanding the intricate relationship between environmental, meteorological, and consumption factors and the adoption of UPACs in Portugal. The study aims to provide actionable insights for policymakers, energy managers, and stakeholders in sustainable energy planning and infrastructure development. The findings can contribute to optimizing renewable energy integration strategies, forecasting sustainable practices, and informing future energy policies.

The choice of the ML models in this study was driven by the need to capture the complex relationships between various factors influencing the adoption of UPACs across municipalities in Portugal. Various supervised ML models were employed for analysis, including:

1. **Linear Regression:** the baseline model in the analysis assumes a linear relationship between the independent variables and the target variables.
2. **Polynomial Regression:** extends the linear model by introducing polynomial terms, allowing for the exploration of non-linear relationships between variables.
3. **Lasso and Ridge Regression:** to address potential multicollinearity among independent variables.
4. **Decision Tree:** a non-linear model that captures complex relationships through a series of binary decisions.
5. **Random Forest:** leverages multiple decision trees to enhance predictive accuracy and generalization, mitigates overfitting, captures intricate patterns within the data, and is particularly robust in handling many variables.
6. **XGBoost:** builds a series of weak learners, optimizing the model's predictive accuracy. XGBoost is well-suited for this study due to its ability to capture non-linear relationships and interactions among variables.

Including a diverse set of ML models ensures a comprehensive exploration of the relationships within the dataset. Each model allows for a nuanced understanding of the factors influencing the adoption of UPAC units in Portugal's energy landscape.

The target variable in these models is the Number of Installations, representing the count of UPAC installations in each municipality. The independent variables include the meteorological variables (Shortwave Radiation, Direct Radiation, Diffuse Radiation, Direct Normal and Irradiance, Maximum Temperature Minimum Temperature, Precipitation), as well as the UPAC Total Installed Power (kW), Energy Consumption (kW), Power Range, and Municipalities.

The chosen metric for the model evaluation is the MSE, which measures the average squared difference between predicted and actual values. MSE is chosen given the real-world implications of this study, where accurately predicting the number of UPAC installations is crucial and for its simplicity and interpretability.

To test the first hypothesis, examining the ML model coefficients provides an understanding of the influence of environmental and meteorological factors on the adoption of UPACs in municipalities. Positive coefficients signify a positive relationship with the Number of Installations, indicating that environmental and meteorological conditions contribute positively to the uptake of UPACs. Whereas negative coefficients signify a negative relationship with the Number of Installations, when the value of the variable increases the Number of Installations decreases.

To test the second hypothesis, statistical significance is evaluated using a t-test with an alpha level of 0.05. This test assesses whether the differences in MSE between models are significant, providing a rigorous statistical framework to validate the hypothesis.

It is important to note that the dataset was separated into training and testing sets in an 80/20 ratio, and cross-validation techniques were employed to enhance the robustness of the model evaluation. A systematic search for the best model parameters was conducted, ensuring the models were fine-tuned for optimal predictive performance.

### 6.5. Data Analysis

The data used in this study was sourced from E-Redes. E-Redes is the company that manages the energy distribution network in Portugal. The data is available at the Open Data Portal from E-Redes and is entitled “Total Production Units for Self-Consumption”. The available variables are Trimester, District, Municipality, Installed Power Range (kW), Number of Installations, and UPAC Total Installed Power (kW).

This dataset spans 2022 to the second trimester of 2023 and encompasses information on installations of UPACs from all municipalities in Portugal. UPAC is an installation to produce electrical energy from renewable energy intended for self-consumption in the installation of associated use, with the possibility of injecting surpluses into the RESP. Individual self-consumers produce renewable energy at their facilities within the national territory, which can be stored or sold. This activity cannot constitute their main commercial or professional activity for non-domestic renewable energy self-consumers.

To prepare the data for the intended analysis, it was necessary to collect additional data, namely the energy consumption in each municipality and meteorological factors. The energy consumption of each municipality is also available at the Open Data Portal from E-Redes, entitled: “Monthly Consumption by Municipality”.

The meteorological variables for understanding the environmental context were collected using an open-source weather API from the Open-Meteo website. This API provided access to a range of meteorological parameters, and the following variables were retained for the study:

1. **Shortwave Radiation; Direct Radiation:** the amount of sunlight or solar radiation, especially direct radiation, is a crucial factor for the efficiency of PV panels. Higher solar radiation levels typically lead to greater electricity generation from PV panels.
2. **Diffuse Radiation:** diffuse radiation can contribute to the overall energy yield of PV panels, especially on cloudy or overcast days.

3. **Direct Normal Irradiance:** this measure of solar radiation is directly related to the energy that PV panels can capture. Panels are most efficient when they receive sunlight at an optimal angle, which is also affected by the sun's angle in the sky.
4. **Maximum Temperature, Minimum Temperature:** Temperature affects the performance of PV panels. High temperatures can reduce the efficiency of PV panels, and their performance may drop in hot weather.
5. **Precipitation:** excessive precipitation, especially over extended periods, can reduce the amount of sunlight that reaches the PV panels.

The collected datasets were submitted for careful preparation to ensure the reliability and consistency of the information used in the analysis. The “Total Production Units for Self-Consumption” dataset is structured in trimesters, representing a unique temporal dimension that needed careful consideration during data preparation. To align with the trimester structure of the original dataset, the “Monthly Consumption by Municipality” dataset was manipulated to aggregate monthly energy consumption into trimester totals. This involved summing the monthly consumption values for each municipality within a trimester. For the meteorological variables, the mean values for each meteorological parameter were calculated for each trimester to capture the broader climatic trends within each trimester, acknowledging the potential impact of seasonal variations on the adoption of UPACs.

In the exploratory phase of the analysis, the focus was deriving essential descriptive statistics for the target variable (Table 1). The mode indicates that the most common count of installations is 1, shedding light on the prevalent occurrence in the dataset. The standard deviation of 338 indicated considerable variability around the mean, highlighting the diverse adoption patterns across municipalities.

Table 1. Descriptive Statistics - Number of Installations.

| | Number of Installations |
|---|---|
| Mean | 120 |
| Mode | 1 |
| Std | 338 |
| Min | 1 |
| 25% | 4 |
| 50% | 11 |
| 75% | 65 |
| Max | 3907 |

To minimize the influence of outliers, the quantile 90% was analyzed. It indicated that 90% of the dataset has 299 or fewer UPAC installations. A histogram (Figure 6) was constructed to visualize this distribution more specifically, considering only data points where the Number of Installations is smaller than or equal to 299.

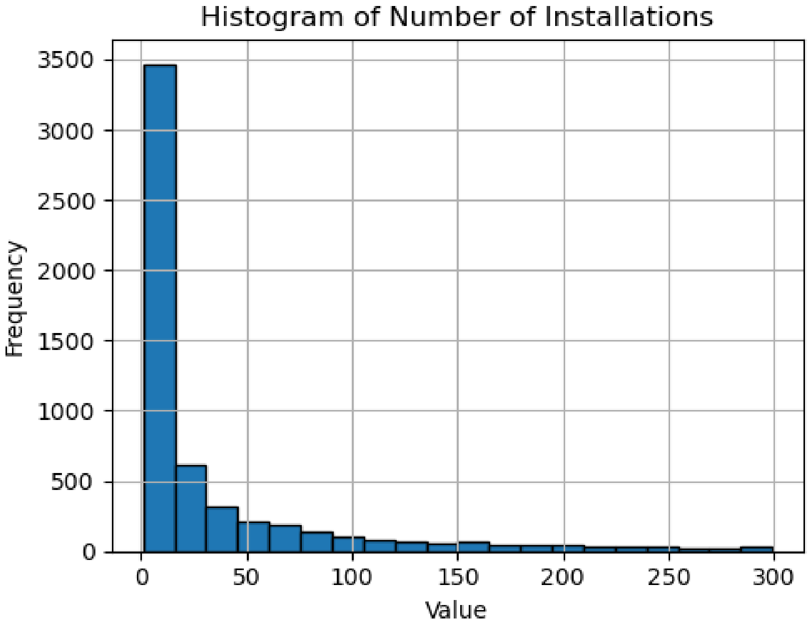


Figure 6. Histogram of Number of Installations.

The "Installed Power Range (kW)" categorizes UPAC installations based on their capacities. Among the reported categories, most installations fall within the 0 to 4 kW power range, as shown in Figure 7. This indicates a prevalent adoption of smaller-scale UPAC units, which are well-suited for individual households or smaller commercial entities.

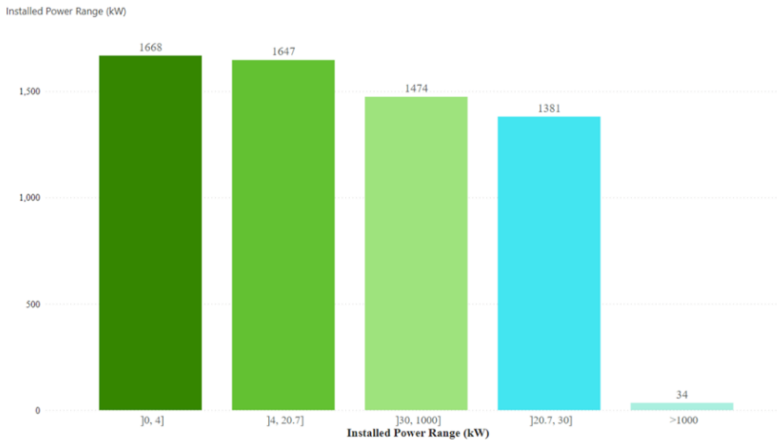


*Figure 7. Installed Power Range (kW).*

As illustrated in Figure 8, the concentration of high UPAC Total Installed Power (kW) values within the Littoral North region is notably pronounced, indicating an emphasis on renewable energy adoption in this geographic area. Municipalities such as Leiria, Guimarães, Braga, and Barcelos exemplify the solid commitment to larger-scale UPAC units, contributing significantly to the regional renewable energy landscape. Additionally, the municipality of Loulé stands out as a notable exception outside the Littoral North.

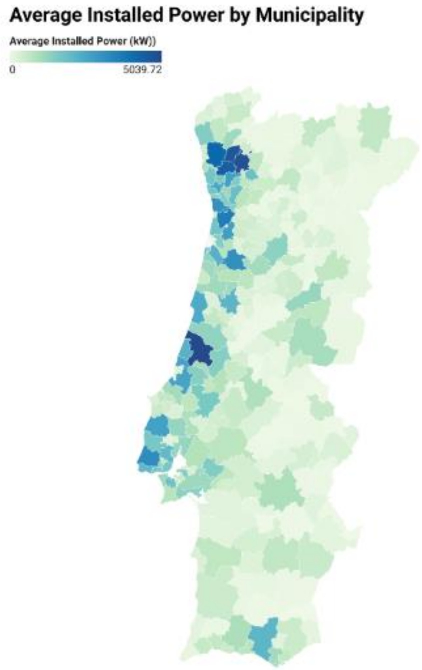


*Figure 8. Average Installed Power by Municipality.*

A correlation matrix (see Appendix A) was constructed to examine the relationships between the various variables and the potential patterns and interactions (see Appendix B). The meteorological variables exhibit strong correlations among themselves. However, the target variable does not strongly correlate with any specific variable (Figure 9). This lack of correlation implies that UPAC adoption across municipalities may be influenced by a combination of factors, with meteorological conditions playing a role alongside other variables such as Power Range and Energy Consumption.

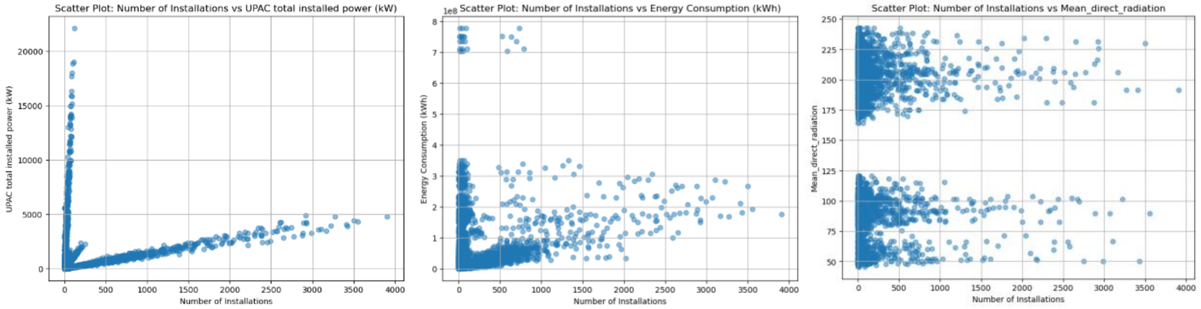


*Figure 9. Scatter Plots Showcasing the Relationship Between the Number of Installations and Key Variables: UPAC Total Installed Power, Energy Consumption, Direct Radiation.*

The study employed a diverse set of ML models to analyze the factors influencing the adoption of UPACs across municipalities in Portugal. The Linear Regression model, serving as a baseline mode, revealed an MSE of 55 673.68, such a high MSE revealed that this linear model has limitations in capturing the complexity within the dataset. The Polynomial Regression model significantly enhances predictive accuracy, achieving a substantially lower MSE of 2 155.85, this improvement signifies the model's capability to accommodate non-linear relationships and underscores its enhanced predictive accuracy. Lasso Regression showcased an exceptional performance, with a low MSE of 259.88; because of Lasso's ability to address multicollinearity among independent variables, the model's predictive performance was strong.

In contrast, Ridge Regression exhibited a higher MSE of 90 813.46, indicating less effectiveness in capturing the underlying patterns in the data. The Decision Tree model performed well, achieving an MSE of 2 598.33, effectively capturing complex relationships through binary decisions. The Random Forest model leveraged multiple decision trees and demonstrated robust performance with an MSE of 1 759.73, its ensemble nature enhanced predictive accuracy and generalization.

XGBoost displayed competitive results, with an MSE of 844.81, the ability of XGBoost to optimize predictive accuracy through a series of weak learners made it effective in capturing the non-linear relationships among the variables. The two models with the best results are Lasso Regression and XGBoost.

Table 2. Evaluation Metrics of the Models.

| | **MSE** | **MAE** | **R-Squared ($R^2$)** |
|---|---:|---:|---:|
| **Linear Regression** | 55 673.68 | 144.54 | 0.44 |
| **Polynomial Regression** | 2 155.85 | 25.83 | 0.98 |
| **Lasso Regression** | 259.88 | 8.23 | 0.99 |
| **Ridge Regression** | 90 813.46 | 159.34 | 0.09 |
| **Decision Tree** | 2 598.33 | 14.37 | 0.97 |
| **Random Forest** | 1 759.73 | 11.55 | 0.98 |
| **XGBoost** | 844.81 | 9.83 | 0.99 |

Examining coefficients (see Attachment Metrics and Coefficients) also provides insights into the strength and direction of relationships in linear and regularized regression models. This information is essential as it can guide policymakers in understanding the key drivers of UPAC adoption. The top five coefficients for the two best models (Lasso Regression; XGBoost) can be examined (Table 3 and Table 4).

Table 3 shows the five most influential coefficients in the Lasso Regression and its intercept. The intercept is the baseline value when all the predictor variables are zero. For this model, the predicted value is -22.63. The most robust coefficient of the Lasso Regression model indicates that in Guimarães, UPAC installations in the 0-4 kW range are associated with a significantly higher number of installations. The negative coefficient for Lisboa indicates a negative relationship. In this context, UPAC installations in the 0-4 kW range negatively influence adoption in Lisbon compared to other regions.

Table 3. Lasso Regression Highest Coefficients.

| Lasso Regression | |
|---|---|
| **Feature** | **Coefficient** |
| Intercept | -22.63 |
| PowerRange_0_4 County_Guimarães | 890.65 |
| PowerRange_0_4 County_Vila Nova de Famalicão | 796.99 |
| PowerRange_0_4 County_Seixal | 671.17 |
| PowerRange_0_4 County_Lisboa | -423.26 |
| PowerRange_0_4 County_Sesimbra | 398.30 |

For the XGBoost model, the independent variables with more importance can be observed in Table 4. UPAC Total Installer Power (kW) holds the highest importance, indicating that this variable has the most substantial impact on predicting the number of UPAC installations. The Power Range 0-4 is the second most important feature, suggesting that installations in this power range contribute significantly to the model’s predictive accuracy. These feature importance values align with the coefficient analysis and emphasize the crucial role of Total Installed Power and the Power Range of 0-4 kW range in influencing UPAC adoption.

Table 4. Independent Variable Importance in the XGBoost Model.

| XGBoost | |
|---|---|
| **Feature** | **Importance** |
| UPAC total installed power (kW) | 0.4807 |
| PowerRange_0_4 | 0.2699 |
| County_Vila Nova de Famalicão | 0.0452 |
| County_Guimarães | 0.0365 |
| County_Seixal | 0.0252 |

Visualizing the predicted vs. actual outcomes through graphical representations also helps assess the model’s accuracy (see Appendix C). The alignment of points along a diagonal line indicates a strong correspondence between predicted and actual values, providing a visual confirmation of the model's performance. In Figure 10, it is possible to observe the comparison between the worst (Linear Regression) and best (Lasso Regression) models.

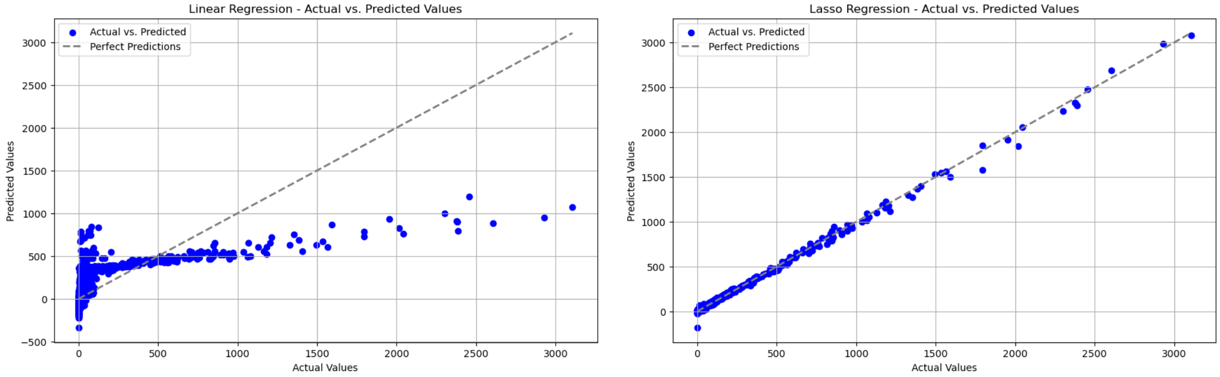


*Figure 10. Comparison of Actual vs. Predicted Values Using Linear and Lasso Regression Models.*

### 6.6. Discussion

The coefficients for meteorological variables in the models and their feature importance support the first hypothesis. These findings emphasize the influence of environmental and meteorological factors on UPAC adoption rates. Shortwave radiation, direct radiation, and direct normal irradiance play fundamental roles, underscoring the importance of sunlight exposure in driving UPAC installations. Additionally, the nuanced impact of meteorological variables is evident in the feature importance values from XGBoost. While Total Installed Power and Power Range remain paramount, meteorological variables contribute substantially to the overall predictive accuracy. This implies that understanding environmental conditions is crucial for accurately predicting UPAC adoption in different municipalities.

The results of the hypothesis testing, comparing the MSE values across different supervised ML models, provide insights into the effectiveness of these models in predicting UPAC adoption. The MSE values, ranging from 259.88 to 90 813.46, reveal varying levels of predictive accuracy. Notably, the MSE of 90 813.46 for the Ridge Regression model suggests a statistically significant difference, indicating potential limitations in capturing the complex relationships within the dataset. The other models, including Lasso Regression, XGBoost, and Random Forest, show MSE values that do not suggest a statistically significant difference from each other, indicating comparable predictive performance. These findings align with the second hypothesis, implying that the selected ML models exhibit similar predictive capabilities when considering the MSE. The chosen models offer consistent and reliable predictions of UPAC unit installations across municipalities in Portugal.

The findings of this study carry significant implications for practical applications, particularly in developing and implementing policies aimed at promoting the adoption of UPACs in Portugal's municipalities. Policymakers can draw valuable insights from the ML models employed in the study, specifically Lasso Regression. This model offers a robust framework for formulating targeted interventions that cater to the unique characteristics of each region. Strategically, the emphasis on Total Installed Power and specific ranges, such as 0-4 kW, guides infrastructure development initiatives. Encouraging larger-scale installations and supporting the growth of UPAC installations can be critical priorities for policymakers seeking to maximize the effectiveness of renewable energy integration.

Moreover, the influence of meteorological variables, including shortwave and direct radiation, suggests that environmental factors play a crucial role in UPAC adoption. Policymakers can integrate these weather-related factors into their planning, ensuring that initiatives align with the climatic conditions of each region. This holistic approach can enhance the efficacy of policy interventions and facilitate a more sustainable energy landscape in Portugal.

Theoretical advancements in renewable energy adoption can benefit from the insights gained in this study. The preference for non-linear models, such as XGBoost and Random Forest, over linear models highlight the inherently complex and non-linear dynamics of the relationship between predictors and UPAC adoption. Theoretical frameworks should evolve to accommodate and reflect these non-linear dynamics, fostering a deeper understanding of the complexities inherent in the adoption process. The coefficients for meteorological variables underscore the need for theoretical models that adopt a holistic perspective. The interplay of environmental, meteorological, and power-related factors should be considered in theoretical frameworks to understand the adoption process comprehensively. The study suggests that theoretical models should be dynamic, capable of continuous learning, and adaptable to emerging trends in UPAC adoption as datasets evolve and new variables become relevant.

## 6.7. Conclusion

The significance of the findings echoes resoundingly for policymakers and energy planners engaged in sculpting Portugal's sustainable energy future. The primary hypothesis, challenging the statistical significance of supervised ML models in predicting UPAC installations, marks the inception of a journey seamlessly weaving together theoretical frameworks and practical implications. Through rigorous analysis, the study navigates the landscape of UPAC adoption in Portuguese municipalities, shedding light on the nuanced relationships that influence the accuracy of predictive models.

The coefficients observed in environmental and meteorological variables underscore these factors' fundamental role in driving UPAC adoption. This insight not only enriches our understanding of renewable energy dynamics but also furnishes actionable information for stakeholders seeking to enhance the integration of UPACs into Portugal's energy infrastructure.

In conclusion, this study contributes to igniting further research of regional variations in UPAC adoption dynamics and understanding how evolving environmental conditions may impact renewable energy integration.

This concluding chapter solidifies this topic as a foundational piece. It offers a navigational guide for policymakers, energy managers, and stakeholders as they navigate the intricate relationships governing Portugal's journey toward a sustainable energy future. The thesis is a testament to the complexity of renewable energy adoption, providing a roadmap that invites continual exploration and adaptation in pursuing a greener and more resilient energy landscape.

## References

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.

Câmara Municipal Seixal. 2021. “Autarquia Aposta Na Energia Fotovoltaica.” December 16, 2021. https://www.cm-seixal.pt/noticia/autarquia-aposta-na-energia-fotovoltaica.

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.

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.

Direção Geral de Energia e Geologia. n.d. “Política energética.” https://www.dgeg.gov.pt/pt/areas-transversais/relacoes-internacionais/politica-energetica/.

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.

Fundo Ambiental. 2022. “Apoio à Concretização de Comunidades de Energia Renovável e Autoconsumo Coletivo.” Recuperar Portugal. https://recuperarportugal.gov.pt/candidatura/02-apoio-a-concretizacao-de-comunidades-de-energia-renovavel-e-autoconsumo-coletivo-aviso-de-abertura-de-concurso-n-o-02-c13-i03-2022/.

Goldenergy. 2022. “Incentivos e benefícios fiscais dos painéis solares.” Goldenergy (blog). March 22, 2022. https://goldenergy.pt/blog/autoconsumo/beneficios-fiscais-paineis-solares/.

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.

Idealista. 2023. “Energia: Portugal aposta nas unidades de produção para autoconsumo.” Idealista/News, April 18, 2023. https://www.idealista.pt/news/financas/economia/2023/04/18/57457-energia-portugal-aposta-nas-unidades-de-producao-para-autoconsumo.

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.

Matura Pragash, Saianiruth & K, Vinoth & P, Karthik. (2023). Smart Grid Technology and Its Impact on Renewable Energy Integration. https://www.researchgate.net/publication/368690043_Smart_Grid_Technology_and_Its_Impact_on_Renewable_Energy_Integration.

Município de Guimarães. 2023. “Guimarães lidera na instalação de Painéis Solares para autoconsumo.” Câmara Municipal De Guimarães. April 10, 2023. https://www.cm-guimaraes.pt/viver/noticias/noticia/guimaraes-lidera-na-instalacao-de-paineis-solares-para-autoconsumo.

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

PORTUGAL.GOV.PT. 2023. “Um país mais verde, mais cedo: o Plano Nacional de Energia e Clima 2030 está a ser revisto pela primeira vez (VÍDEO).” June 30, 2023. https://www.portugal.gov.pt/pt/gc23/comunicacao/noticia?i=-um-pais-mais-verde-mais-cedo-o-plano-nacional-de-energia-e-clima-2030-esta-a-ser-revisto-pela-primeira-vez.

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.

Ribeiro, Sara. 2022. “Capacidade instalada para autoconsumo de energia mais que duplicou este ano.” Dinheiro Vivo, September 4, 2022. https://www.dinheirovivo.pt/economia/capacidade-instalada-para-autoconsumo-de-energia-mais-que-duplicou-este-ano-15135716.html.

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.

Silva, Pedro, and Antonieta Duarte. 2023. “Fundo ambiental: como concorrer ao Programa de Apoio a Edifícios mais Sustentáveis 2023.” Edited by Ana Rita Costa and Alda Mota. DECO PROTESTE. September 20, 2023. https://www.deco.proteste.pt/casa-energia/aquecimento/noticias/fundo-ambiental-como-concorrer-programa-apoio-edificios-mais-sustentaveis-2023#em-que-consiste-o-programa-de-apoio-a-edificios-mais-sustentaveis.

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/.

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.

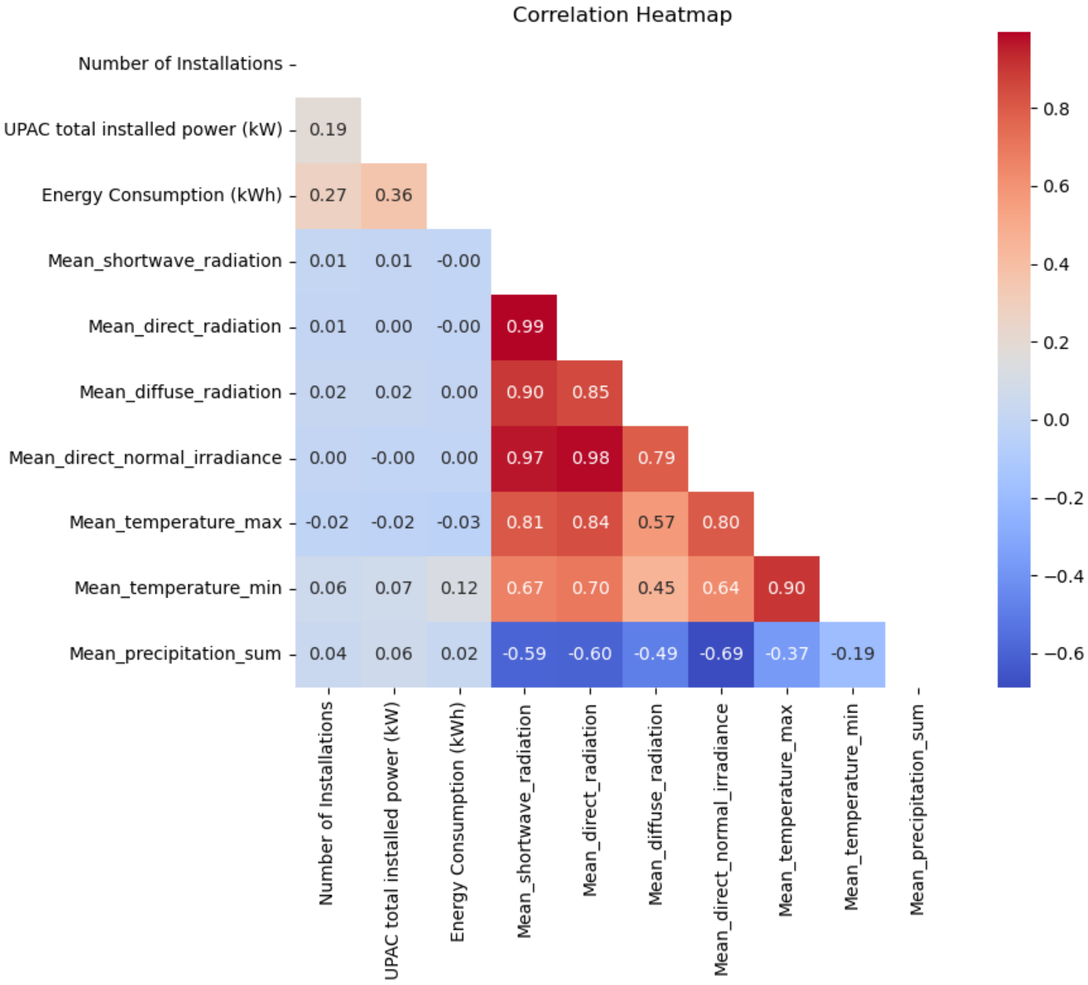
Zhang, Qi, Hailong Li, Lei Zhu, Pietro Elia Campana, Huihui Lu, Fredrik Wallin, and Qiang Sun. 2018. “Factors Influencing the Economics of Public Charging Infrastructures for EV – A Review.” Renewable & Sustainable Energy Review

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

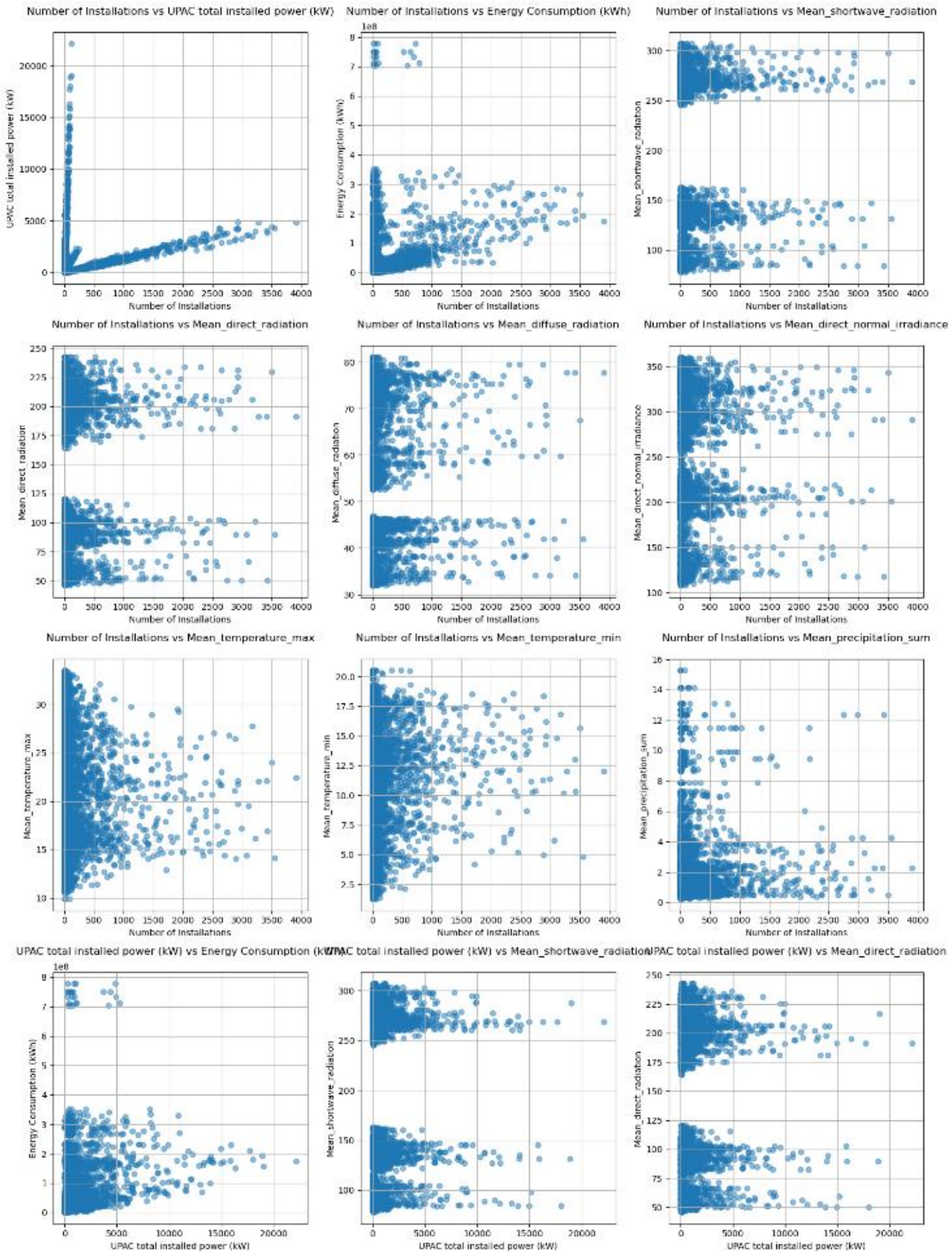
# Appendix

Appendix A. Correlation Matrix.

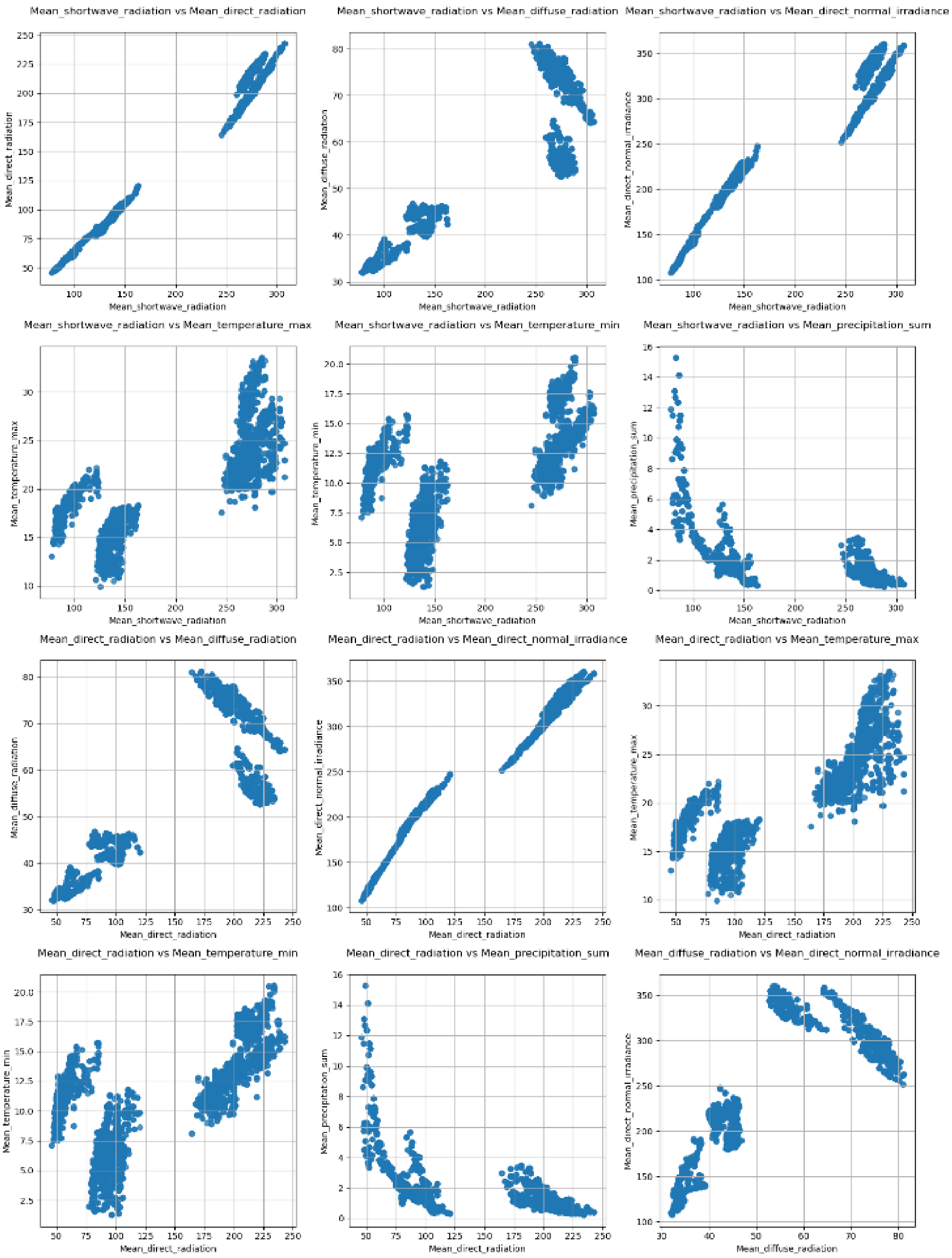


| | Number of Installations | UPAC total installed power (kW) | Energy Consumption (kWh) | Mean_shortwave_radiation | Mean_direct_radiation | Mean_diffuse_radiation | Mean_direct_normal_irradiance | Mean_temperature_max | Mean_temperature_min | Mean_precipitation_sum |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of Installations | | | | | | | | | | |
| UPAC total installed power (kW) | 0.19 | | | | | | | | | |
| Energy Consumption (kWh) | 0.27 | 0.36 | | | | | | | | |
| Mean_shortwave_radiation | 0.01 | 0.01 | -0.00 | | | | | | | |
| Mean_direct_radiation | 0.01 | 0.00 | -0.00 | 0.99 | | | | | | |
| Mean_diffuse_radiation | 0.02 | 0.02 | 0.00 | 0.90 | 0.85 | | | | | |
| Mean_direct_normal_irradiance | 0.00 | -0.00 | 0.00 | 0.97 | 0.98 | 0.79 | | | | |
| Mean_temperature_max | -0.02 | -0.02 | -0.03 | 0.81 | 0.84 | 0.57 | 0.80 | | | |
| Mean_temperature_min | 0.06 | 0.07 | 0.12 | 0.67 | 0.70 | 0.45 | 0.64 | 0.90 | | |
| Mean_precipitation_sum | 0.04 | 0.06 | 0.02 | -0.59 | -0.60 | -0.49 | -0.69 | -0.37 | -0.19 | |

*Appendix B. Interaction Between Variables.*




UPAC total installed power (kW) vs Mean_diffuse_radiation
Mean_diffuse_radiation
UPAC total installed power (kW)
UPAC total installed power (kW) vs Mean_direct_normal_irradiance
Mean_direct_normal_irradiance
UPAC total installed power (kW)
UPAC total installed power (kW) vs Mean_temperature_max
Mean_temperature_max
UPAC total installed power (kW)
UPAC total installed power (kW) vs Mean_temperature_min
Mean_temperature_min
UPAC total installed power (kW)
UPAC total installed power (kW) vs Mean_precipitation_sum
Mean_precipitation_sum
UPAC total installed power (kW)
Energy Consumption (kWh) vs Mean_shortwave_radiation
Mean_shortwave_radiation
Energy Consumption (kWh)
Energy Consumption (kWh) vs Mean_direct_radiation
Mean_direct_radiation
Energy Consumption (kWh)
Energy Consumption (kWh) vs Mean_diffuse_radiation
Mean_diffuse_radiation
Energy Consumption (kWh)
Energy Consumption (kWh) vs Mean_direct_normal_irradiance
Mean_direct_normal_irradiance
Energy Consumption (kWh)
Energy Consumption (kWh) vs Mean_temperature_max
Mean_temperature_max
Energy Consumption (kWh)
Energy Consumption (kWh) vs Mean_temperature_min
Mean_temperature_min
Energy Consumption (kWh)
Energy Consumption (kWh) vs Mean_precipitation_sum
Mean_precipitation_sum
Energy Consumption (kWh)
1e8
0
5000
10000
15000
20000


Mean_shortwave_radiation vs Mean_direct_radiation

Mean_shortwave_radiation vs Mean_diffuse_radiation

Mean_shortwave_radiation vs Mean_direct_normal_irradiance

Mean_shortwave_radiation vs Mean_temperature_max

Mean_shortwave_radiation vs Mean_temperature_min

Mean_shortwave_radiation vs Mean_precipitation_sum

Mean_direct_radiation vs Mean_diffuse_radiation

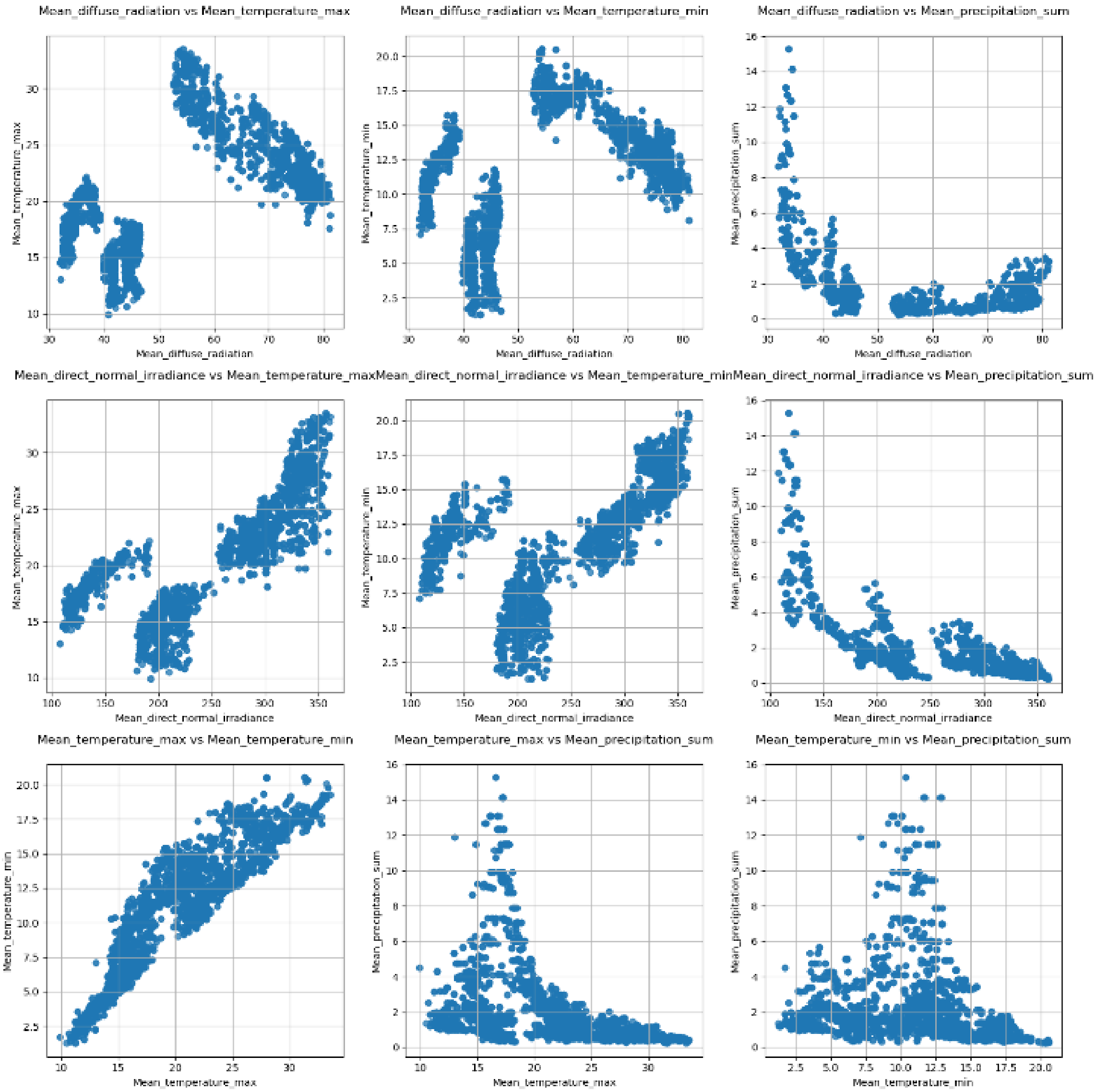
Mean_direct_radiation vs Mean_direct_normal_irradiance

Mean_direct_radiation vs Mean_temperature_max

Mean_direct_radiation vs Mean_temperature_min

Mean_direct_radiation vs Mean_precipitation_sum

Mean_diffuse_radiation vs Mean_direct_normal_irradiance


Mean_diffuse_radiation vs Mean_temperature_max
Mean_diffuse_radiation vs Mean_temperature_min
Mean_diffuse_radiation vs Mean_precipitation_sum
Mean_direct_normal_irradiance vs Mean_temperature_max
Mean_direct_normal_irradiance vs Mean_temperature_min
Mean_direct_normal_irradiance vs Mean_precipitation_sum
Mean_temperature_max vs Mean_temperature_min
Mean_temperature_max vs Mean_precipitation_sum
Mean_temperature_min vs Mean_precipitation_sum
Mean_diffuse_radiation
Mean_direct_normal_irradiance
Mean_temperature_max
Mean_temperature_min
Mean_precipitation_sum
30
40
50
60
70
80
100
150
200
250
300
350
10
15
20
25
30
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
0
2
4
6
8
10
12
14
16

## *Appendix C. Actual vs. Predict Values.*