INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

MESTRADO EM ENGENHARIA INFORMÁTICA

Electricity Price Forecasting

GONÇALO MIGUEL CARVALHO PEREIRA

Junho de 2025

POLITÉCNICO
DO PORTO

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

Electricity Price Forecasting

Gonçalo Miguel Carvalho Pereira

A dissertation preparation for the degree of Master of Science, Specialisation Area of Information and Knowledge Systems

Advisor: Fátima Rodrigues

Porto, June 26, 2025

Statement of Integrity

I hereby declare having conducted this academic work with integrity.

I have not plagiarised or applied any form of undue use of information or falsification of results along the process leading to its elaboration.

Therefore the work presented in this document is original and authored by me, having not previously been used for any other end.

I further declare that I have fully acknowledged the Code of Ethical Conduct of P.PORTO.

ISEP, Porto, June 26, 2025

Dedicatory

This dissertation marks the end of a journey filled with learning, hard work, and personal growth. It represents months of effort and resilience, along with moments of doubt, but also achievement and pride.

First, I want to thank my family for their unwavering support and encouragement during this journey. Their belief in me, especially during tough times, was a constant source of strength.

To my girlfriend, Maria, thank you for your patience and motivation. Your support made difficult steps easier and each small victory more meaningful.

To my friends who walked this path with me in different ways, thank you for your encouraging words, for helping me take breaks when needed, and for celebrating each milestone with me.

A special thanks to Professor Fátima Rodrigues. Her guidance and experience were essential throughout this process. Her helpful feedback and support were key to making this research something I am proud of.

Finally, I dedicate this work to everyone who believed in me, even when I doubted myself. This is also yours.

Abstract

This dissertation examines the application of advanced machine learning techniques to enhance electricity price forecasting in deregulated markets. As electricity markets continue to evolve with deregulation and increased competition, accurate price forecasting becomes vital for stakeholders, including energy producers, grid operators, and traders. Traditional forecasting models often face challenges in capturing the complex, non-linear dynamics of electricity prices, which are influenced by various factors such as weather patterns, demand fluctuations, supply-side uncertainties, and regulatory changes.

This research employs the CRISP-DM methodology, utilizing machine learning algorithms such as Long Short-Term Memory (LSTM), Random Forest, and XGBoost to improve both short-term and long-term price predictions. The methodology involves data collection and preprocessing, model development, variable tuning, and validation against real-world data. Key findings indicate that integrating multiple data sources, including real-time weather data and demand forecasts, significantly enhances the accuracy and reliability of forecasting models.

Throughout this work, several forecasting models, both statistical and machine learning-based, were developed and compared for short-term and long-term horizons. The best short-term results were achieved using LSTM and XGBoost, with MAPE values below 8% after hyperparameter tuning. For long-term forecasting, XGBoost and Random Forest stood out for their robustness and stability, with XGBoost achieving a MAPE of 9.78%. Statistical tests confirmed that in certain contexts, the performance differences between models were statistically significant, thus validating the methodological choices made.

Keywords: Electricity Price Forecasting, Machine Learning, Deregulated Markets, Forecasting Models, Data Analysis

Resumo

Esta dissertação explora a aplicação de técnicas avançadas de machine learning para melhorar a previsão de preços de eletricidade em mercados desregulados. À medida que os mercados de eletricidade continuam a evoluir com a desregulação e o aumento da concorrência, a previsão precisa dos preços torna-se vital para os diferentes intervenientes, incluindo produtores de energia, operadores de rede e traders. Os modelos tradicionais de previsão enfrentam frequentemente dificuldades em capturar as dinâmicas complexas e não lineares dos preços da eletricidade, que são influenciadas por diversos fatores como padrões meteorológicos, flutuações na procura, incertezas do lado da oferta e alterações regulatórias.

Esta investigação segue a metodologia CRISP-DM, utilizando algoritmos de machine learning como Long Short-Term Memory (LSTM), Random Forest e XGBoost para melhorar as previsões de preços a curto e a longo prazo. A metodologia abrange a recolha e pré-processamento de dados, o desenvolvimento dos modelos, a afinação de variáveis e a validação com dados reais. Os principais resultados indicam que a integração de múltiplas fontes de dados, incluindo dados meteorológicos em tempo real e previsões de procura, melhora significativamente a precisão e fiabilidade dos modelos de previsão.

Ao longo deste trabalho foram desenvolvidos e comparados vários modelos de previsão, tanto estatísticos como de machine learning, para horizontes de curto e longo prazo. Os melhores resultados no curto prazo foram obtidos com os modelos LSTM e XGBoost, atingindo um erro MAPE inferior a 8% após afinação. Para o longo prazo, os modelos XGBoost e Random Forest destacaram-se pela sua robustez e estabilidade, com o XGBoost a alcançar um MAPE de 9.78%. Os testes estatísticos confirmaram que, em determinados contextos, as diferenças entre modelos são estatisticamente significativas, validando as escolhas metodológicas efetuadas.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and Motivation

Electricity markets have been changing over the last decades, with the majority of countries adopting deregulation and liberalization frameworks. In a deregulated electricity market, the forces of supply and demand determine the price rather than a central authority. This shift has brought more competition, offering consumers wider choices and enabling efficient energy trading. With deregulation, however, comes the complication of being able to forecast electricity prices because of such variables as changing demand, uncertainty on the supply side, and weather conditions.

Such a market would also require accurate price forecasting for various stakeholders with regard to electricity. Energy producers, grid operators, and traders make real-time decisions concerning the generation and consumption of energy based on precise short-term forecasts. In contrast, long-term forecasts are important in helping formulate investment strategies, regulate planning, and manage risk. The ability to forecast both short and long-term electricity prices accurately becomes all-important in order to ensure that operations are optimized along with stability in the energy systems.

Recent developments in machine learning have shown great potential in solving complex forecasting problems. Traditional models, such as ARIMA, have been used to predict time series data but often fall short when dealing with the highly nonlinear and dynamic nature of electricity prices. Newer approaches, like LSTM networks and Random Forest, could allow a higher level of accuracy in forecasts by considering hundreds of years of historical data and real-time data.

This dissertation explores improving electricity price forecasting in deregulated markets with advanced machine learning techniques. The goal will be to devise robust models that give reliable short and long-term predictions which would help market participants optimize their decision-making processes.

1.2 Problem Statement

Deregulated electricity markets encounter a multitude of challenges in the realm of price forecasting. Prices in these markets are influenced by a diverse array of factors, including:

- Weather patterns, which affect both supply, through renewable energy sources like wind and solar, and demand, through temperature changes.

- Demand fluctuations, caused by economic activity, consumer behavior, and industrial needs.
- Supply-side uncertainties, particularly due to the growing penetration of renewable energy sources, which are inherently variable.
- Regulatory changes and economic indicators, which can significantly impact long-term price trends.

Current forecasting models often fail to adequately capture the complexities of these markets, which are typically nonlinear and exhibit high volatility. Consequently, there exists a pressing need for the development of advanced forecasting models that account for these multifaceted variables. Such models would facilitate more accurate and reliable price forecasts over both short-term intervals, including hourly or daily timeframes, and long-term periods, such as weekly, monthly, or yearly horizons.

1.3 Research Objectives

The main objective of this dissertation is to enhance electricity price forecasting in deregulated markets using advanced machine learning models. The specific goals are as follows:

- Comparison of Machine Learning Algorithms: Evaluate and compare the performance of various machine learning algorithms (e.g., ARIMA, LSTM, Random Forest, Gradient Boosting) for both short-term and long-term price forecasting. This comparison will focus on metrics such as accuracy, robustness, and computational efficiency.
- Feature Selection and Engineering: Identify and extract relevant features from historical data to improve the accuracy of the models. This will include factors like weather patterns, demand, and supply-side variables.
- Data Integration: Incorporate multiple data sources, including real-time weather data, demand-side management systems, and supply-side forecasts, to enhance the reliability of the forecasting models.
- Validation with Real-World Data: Validate the proposed models using real-world data from deregulated markets such as the European Energy Exchange (EEX) or the New York Independent System Operator (NYISO). The impact of the developed models on energy trading and risk management will be assessed.

1.4 Research Questions

To achieve the outlined objectives, the following research questions will be addressed:

- How does the forecasting horizon (short-term vs. long-term) affect the performance and suitability of different machine learning models for electricity price prediction?
- Which machine learning algorithms provide the most accurate and robust forecasts for electricity prices, considering factors such as economic trends and regulatory changes?
- What are the most relevant features that influence electricity price movements, and how can they be extracted from historical and real-time data?

1.5 Methodology Overview

This research will be using the CRISP-DM methodology (Wirth and Hipp 2000), one of the more common methodologies in data science that provides a structured framework for tackling data-driven problems. It includes six major phases:

- Data Collection and Preprocessing: Historical data on electricity prices, demand, supply, weather, and other relevant factors will be collected from a dataset found in (Roussis 2021). The data will be preprocessed, with missing values handled, outliers removed, and relevant features engineered.
- Feature Engineering: Feature selection and engineering will be performed to identify the most important variables affecting price fluctuations, and new features will be created to improve model accuracy.
- Model Development: Various machine learning algorithms will be used for both short-term and long-term forecasting. Traditional models such as ARIMA will be compared with more advanced techniques like LSTM, Random Forest, and Gradient Boosting.
- Model Evaluation: The models will be evaluated based on their performance metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Relative MAE (rMAE).
- Deployment: This phase involves integrating the final model into a real-world system for operational use. However, deployment is not performed in this dissertation, as the focus is on model development and evaluation rather than real-time implementation.

1.6 Ethical Considerations

This dissertation follows the conventional fundamental ethical principles that have been accepted, such as the ACM Code of Ethics, IEEE Code of Ethics, and the NSPE Code of Ethics. Such guidelines are supported by three cornerstones of research processes: integrity, openness, and enhancement of societal well-being.

The ethical treatment of data is carried out with due care. By nature, the data of electricity markets is sensitive or proprietary; it is handled in conformance with prescribed regulations, such as the GDPR. Data anonymization and storage procedures ensure that any data used is strictly for the purposes of this study and respects the privacy and confidentiality of all parties.

Recognizing the risks of bias within machine learning models, this research is informed by strategies promoting model fairness and representativeness. These include the development of diverse datasets, in-depth fairness audits, and the deployment of bias mitigation techniques as part of the production of forecasting models that do not inadvertently disadvantage specific participants in the market or widen inequities.

It is therefore methodological in approach, with interpretability in the results. In such forecasting models, an attempt is made to increase the interpretability by incorporating into them various tools and techniques that would allow the stakeholder to understand how a particular prediction is made. All assumptions made, key processes involved, or any limitations are duly recorded; hence accountability and trust are taken into consideration.

Finally, there are the social and environmental impacts of this work considered throughout the research process. The study shall contribute to efficiency and stability in the electricity market, integration of renewable energy sources, and reduction in price volatility for the ultimate benefit of consumers, and further contribute toward sustainable energy goals. Attention is made to ensure that none of the outcomes inadvertently facilitates unethical practices, like market manipulation or inequitable pricing.

Correspondingly, this dissertation addresses these dimensions in the interests of balancing innovation with responsibility by ensuring that the research is useful to both the academe and society at large.

Additionally, selected productivity-enhancing tools based on artificial intelligence were employed throughout the project to streamline routine tasks, such as generating parameter combinations for testing or assisting in the organization of data summaries and tables. These were used strictly to support the analytical workflow, ensuring efficiency without compromising the integrity or originality of the research.

1.7 Structure of the Dissertation

The structure of this dissertation is organized as follows:

- Chapter 2 – Literature Review: This chapter provides an overview of the existing literature on electricity price forecasting in deregulated markets, covering both traditional statistical methods and advanced machine learning approaches.
- Chapter 3 – Data Selection and Preprocessing: This chapter describes the criteria for selecting the dataset, followed by the data cleaning and preprocessing steps applied to prepare the input features for model training.
- Chapter 4 – Exploratory Data Analysis: This chapter presents a detailed exploratory analysis of the dataset, highlighting key trends, patterns, correlations, and variable distributions relevant to electricity price dynamics.
- Chapter 5 – Model Development and Forecasting: This chapter outlines the implementation and evaluation of various forecasting models for both short-term and long-term electricity price prediction, including hyperparameter tuning and performance comparison.
- Chapter 6 – Conclusion: The final chapter summarizes the main findings, revisits the research questions, and discusses potential directions for future work.

Chapter 2

Literature review

PRISMA (PRISMA 2024) outlines a methodology comprising four principal steps: Identification, Screening, Eligibility, and Inclusion. Each of these steps employs a structured and reproducible approach to filter relevant studies. Relevance was determined based on alignment with the dissertation's objectives, particularly focusing on electricity price forecasting in deregulated markets through machine learning. To assess quality and credibility, priority was given to peer-reviewed journal articles published between 2020 and 2024, especially those employing empirical methodologies and reporting performance metrics (such as MAE and RMSE). Articles were also evaluated for the clarity of their methodology and the inclusion of datasets or real-world validation. This rigorous approach ensured that only studies that were methodologically sound and contextually pertinent were included in the final literature set.

During the Identification phase, researchers compile a comprehensive list of potential studies by searching various databases, registries, and other sources to collect all possibly relevant articles. In the Screening phase, duplicates are eliminated, and a preliminary review of titles and abstracts is conducted, excluding studies that do not meet the set inclusion criteria. The Eligibility phase involves an in-depth review of the full texts of the remaining studies against predefined criteria, including study design, population, interventions, and outcomes. Finally, the Inclusion phase narrows down the selection to a final set of studies that fulfill all necessary requirements for either the systematic review or meta-analysis. Each step guarantees a comprehensive and transparent process with minimal bias in selecting highly relevant and methodologically robust studies.

2.1 PRISMA Methodology Overview

The PRISMA methodology is comprised of four pivotal steps: Identification, Screening, Eligibility, and Inclusion. Each of these stages involves a systematic approach to the evaluation and selection of literature, aimed at meticulously eliminating studies that do not meet predefined criteria. This rigorous process is essential to ensure that the final compilation consists solely of highly relevant research articles for comprehensive review.

2.1.1 Identification

In this step, a systematic search was conducted for studies eligible for inclusion in the review across various databases. The primary database utilized was the B-on database, which serves as a central hub providing access to an extensive range of scholarly articles and research publications. By conducting searches within B-on, we ensured comprehensive coverage of the available literature, thereby maximizing the likelihood of identifying all studies pertinent to the research objectives.

The following keywords were used to query the database:

- "Electricity price forecasting"
- "Machine learning electricity market"
- "Short-term electricity price prediction"
- "Long-term electricity price forecasting"
- "Deregulated electricity market"
- "Deep learning"
- "ARIMA, LSTM, Random Forest, Gradient Boosting forecasting"

The keywords were combined with appropriate Boolean operators to render the search comprehensive. Further, filtering was added to relevance by ensuring that only works published in 2020 or later were returned. That way, current progress and developments made in both the fields of machine learning and deregulated markets are captured. The approach is balanced: while being comprehensive, it is highly targeted toward the most updated and relevant research.

2.1.2 Screening

Following the initial identification, a title and abstract screening of the articles was conducted. In this regard, selection became refined through the systematic application of predefined inclusion and exclusion criteria. The inclusion criteria consisted of studies that had a direct relationship with the objectives of the research, those which were methodologically sound, and those that fell within the scope of the review. Conversely, the exclusion criteria removed those articles that were irrelevant, or that would not provide sufficient detail for the standard of quality for subsequent analyses, and then allowed a further selection of materials most suitable, representative, and meaningful to undertake this review.

- Inclusion Criteria
  - Studies that focus on short-term or long-term electricity price forecasting.
  - Articles that utilize machine learning techniques such as ARIMA, LSTM, Random Forest, or Gradient Boosting.
  - Studies conducted in deregulated electricity markets.
  - Peer-reviewed articles published between 2020 and 2024.
- Exclusion Criteria
  - Studies that focus solely on regulated electricity markets.
  - Articles that do not apply machine learning models for forecasting.
  - Studies focused on other energy commodities, such as oil or gas.
  - Non-peer-reviewed articles (e.g., blog posts, white papers).

2.1.3 Eligibility

The remaining abstracts were screened for relevance regarding the study aims during this stage. Studies not offering substantial data, methodologies, or information about machine learning-based forecasting in deregulated electricity markets were excluded. The ones presenting related data, detailed algorithms, or comparative analysis of the used techniques were retained for further consideration. The studies identified were included in this process only if they represented meaningful contributions to the topic, otherwise, selection into further review was ensured.

2.1.4 Inclusion

The final selection for critical analysis was made from the pool of selected articles by careful choice, which is strongly aligned with the objectives of the study. Detailed reviews have been carried out in these selected studies to extract relevant aspects, including methodologies adopted in the studies, machine learning algorithms applied, respective applications for forecasting in the deregulated electricity market, datasets and variables used in various analyses, and comparative performance evaluations or any unique insight presented. Detailed exploration of this nature allowed for deep comprehension of contributions and findings from each selected study, which later helped in forming a really sound synthesis of the literature.

- The machine learning models used for forecasting.
- The datasets employed and their sources (e.g., historical market data, weather data, demand-side data).
- The forecasting horizons considered (short-term or long-term).
- The performance metrics used to evaluate model accuracy (e.g., MAE, RMSE).
- Comparisons between different machine learning models and traditional models.

2.2 PRISMA Flow Diagram

A PRISMA Flow Diagram is attached for clarity and visualization of the selection process. It reflects the number of studies identified during the initial search, those screened for relevance, and the final set of studies included in the review. It allows a transparent overview of each step in the process, highlighting the reasons for exclusion and ensuring clarity and reproducibility in the study selection methodology.

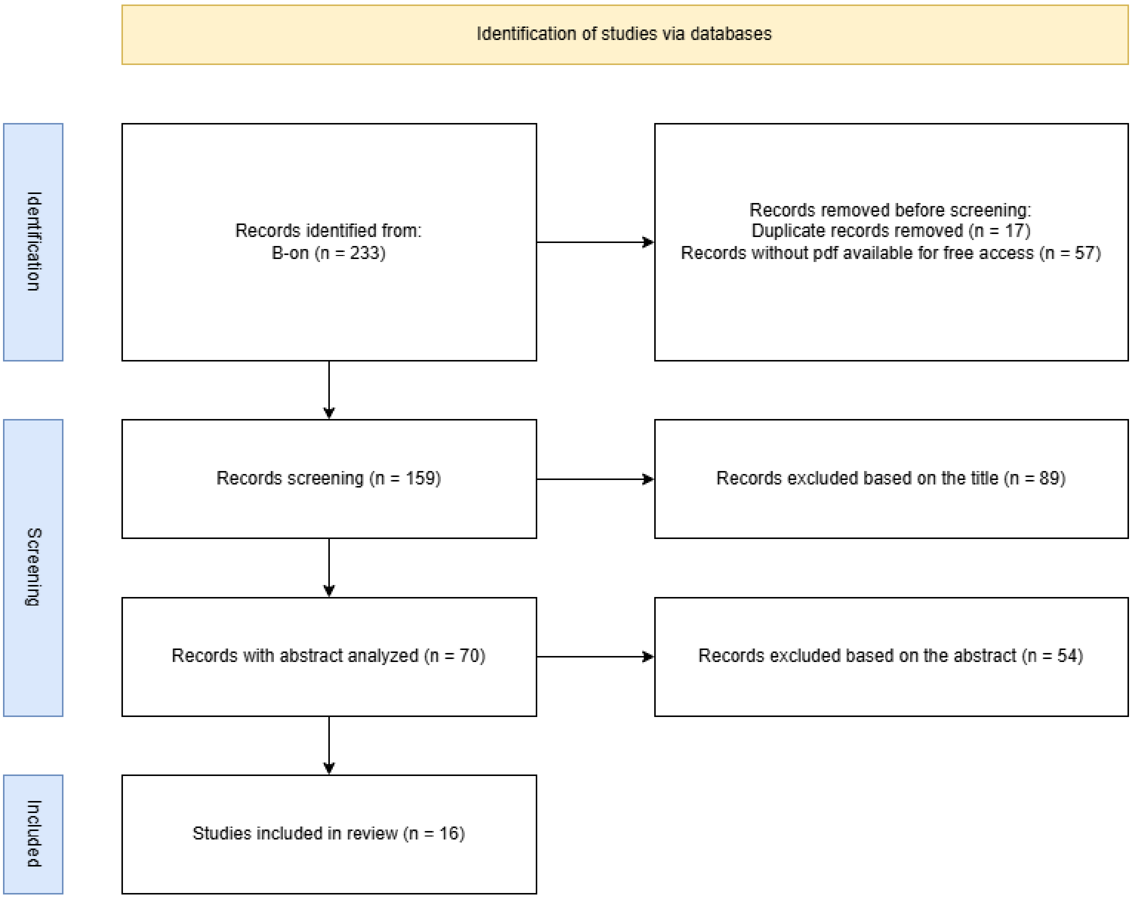

Figure 2.1: PRISMA flow diagram

2.3 Results of the Literature Review

This section covers the key findings and results from the articles studied in this review. These have been structured and organized to present important trends, methodologies, and results concerning machine learning forecasting in deregulated electricity markets. This includes synthesizing identified algorithms and models, including their performance metrics and comparisons against alternative approaches. Furthermore, the conclusion of each study is discussed, highlighting its contribution to the development of the paper on challenges, innovations, and future directions in the field. This critical analysis gives a firm basis for identifying the gaps in literature and areas that potentially warrant further research.

2.3.1 Introduction to Electricity Price Forecasting

Electricity price forecasting is a crucial area of research, as highlighted by (Lago et al. 2021). This field aims to predict both spot and forward prices across various wholesale electricity markets worldwide, each operating under distinct trading mechanisms. For example, in Europe, a significant amount of short-term electricity trading takes place in day-ahead markets, which utilize a uniform-price auction model conducted daily. In contrast, the Australian National Electricity Market (NEM) operates as a real-time power pool, setting dispatch prices every five minutes and calculating average pool prices on a half-hourly basis.

Moreover, (Magalhães et al. 2024) contend that the energy landscape is experiencing a significant transformation, with the traditional role of passive energy consumers evolving into that of active participants in energy management. This shift involves the conversion of "inelastic consumers" into "flexible prosumers" through the integration of advanced data analytics, photovoltaic systems, and modular energy storage technologies. Such innovations empower consumers to adjust their consumption patterns based on off-peak pricing or self-generated electricity. This trend is particularly evident in universities and public services that manage large electrical loads. As a result, the precision of short-term load forecasting (STLF) becomes a vital element in this new framework, enabling optimized load management strategies and informed decisions regarding energy consumption and contractual agreements.

As highlighted by (Bara and Oprea 2024), the current transition toward cleaner and more sustainable energy systems poses significant challenges for electricity price forecasting (EPF). One of the primary concerns is the increased price volatility stemming from the greater integration of renewable energy sources (RES). This volatility is further compounded by unpredictable events, such as the financial crisis of 2008 and geopolitical disruptions. Additionally, as discussed by (Tangtang Wang and Xu 2024), traditional forecasting methodologies—such as Autoregressive Integrated Moving Average (ARIMA) and Autoregressive (AR) models—often fall short in addressing the complexities of contemporary electricity markets. These models struggle to incorporate essential variables, including carbon dioxide (CO2) certificate prices, renewable energy feed-ins, and external market influences. Furthermore, the prevalent modeling approaches that rely on short time intervals exhibit limited robustness, thereby restricting their applicability in the dynamic and often turbulent conditions characterizing modern power markets.

2.3.2 Traditional Forecasting Models

ARIMA (Autoregressive Integrated Moving Average) models play a crucial role in time series analysis due to their straightforward structure and effectiveness in modeling linear, stationary data. However, their dependence on stationarity—usually achieved through the process of differencing—and their sensitivity to non-stationary patterns impose certain limitations, especially when addressing complex and volatile series like electricity price data, as highlighted by (Ewa Chodakowska and Nazarko 2021). In empirical studies related to electricity, ARIMA models frequently serve as baseline models; however, they often struggle to capture abrupt fluctuations and the seasonality that characterize daily and hourly market data.

The seasonal variant known as SARIMA (Seasonal ARIMA) enhances the traditional ARIMA model by explicitly incorporating periodic components, such as daily and weekly seasonality, through seasonal parameters (P, D, Q) that relate to a specific seasonal period (s). This model often demonstrates superior performance relative to ARIMA, achieving lower RMSE in time series that exhibit distinct seasonal patterns, particularly in electricity forecasting tasks, as highlighted by (Kamil Szostek and Kusznier 2024). However, its effectiveness is dependent on the characteristics of the data; as illustrated by (Kamil Szostek and Kusznier 2024), SARIMA provided superior forecasts for one wind farm dataset while underperforming compared to the simpler ARIMA model in another scenario. This underscores the critical importance of careful selection of seasonal parameters. In summary, both ARIMA and SARIMA serve as valuable benchmark models, offering interpretability and computational efficiency. However, their limitations in addressing non-linearity and multivariate interactions suggest a need for further investigation into machine learning and deep learning methodologies.

2.3.3 Machine Learning Techniques

Multiple Linear Regression (MLR) is often utilized as a foundational method in electricity forecasting due to its straightforward nature and ease of interpretation. However, the inherent assumption of linearity in MLR limits its capacity to adequately capture the non-linear dependencies and volatility present in electricity markets. As highlighted by (Nti et al. 2020), MLR tends to underperform when applied to real-world electricity datasets that include variables such as weather, demand, and generation.

Random Forest (RF) is a highly regarded ensemble learning technique that has gained significant traction in forecasting electricity prices and loads. Its popularity arises from its ability to capture complex, nonlinear relationships without imposing strict assumptions about data distribution. The RF method constructs multiple decision trees using bootstrapped samples and averages their predictions, effectively reducing variance and improving generalization, as explained by (J. Zhang 2022).

One of the primary advantages of Random Forest (RF) is its resilience against overfitting, particularly in high-dimensional datasets. Additionally, RF excels in managing mixed-type data and handling missing values. As noted by (J. Zhang 2022), this makes it particularly well-suited for analyzing electricity market data, where various correlated variables—such as demand, temperature, and generation type—interact in complex and often unpredictable ways. Recent applications in power system forecasting have demonstrated the method's stability and strong performance, frequently outperforming traditional models when sufficient training data is available.

Crucial hyperparameters in RF include:

- n_estimators (number of trees): increasing this generally reduces variance, though with diminishing returns and increased computation.
- max_depth (maximum tree depth): controls overfitting by limiting how deep trees can grow.
- min_samples_split and min_samples_leaf: ensure splits are meaningful by requiring a minimum number of samples, thus regularizing the tree structure.

A recent study conducted by (Kushan Sandunil 2024) delves into the optimization of power systems and presents compelling evidence that the application of grid-search optimization for the parameters n_estimators, max_depth, min_samples_split, and min_samples_leaf results in significant enhancements in both accuracy and stability when analyzing electrical system data.

Gradient Boosting, implemented through XGBoost, distinguishes itself by building trees sequentially. Each newly added tree aims to minimize the residual errors of the current ensemble, allowing the model to focus on the most challenging observations. This boosting technique typically results in superior predictive accuracy when compared to bagging methods like Random Forest. XGBoost, in particular, enhances its performance by incorporating additional regularization terms into its loss function and facilitating parallel processing. It has become a preferred model for structured data in forecasting competitions, as highlighted in (Tightiz 2024). Its capacity to manage multicollinearity, handle sparsity, and model complex feature interactions makes it an excellent choice for electricity price forecasting, where nonlinearities, regime changes, and exogenous shocks are prevalent. Research indicates that XGBoost often matches or even surpasses the performance of deep learning techniques on tabular energy datasets, especially when supported by effective feature engineering and moderate hyperparameter tuning.

The key hyperparameters include:

- n_estimators (number of boosting rounds)
- max_depth (controls model complexity)
- learning_rate (shrinkage factor that balances learning speed and overfitting)

(Verma 2024) underscores that reducing learning_rate while increasing n_estimators leads to smoother convergence and better generalization.

2.3.4 Deep Learning Approaches

Recurrent Neural Networks (RNNs) have been widely examined in the context of time series forecasting due to their capacity to preserve temporal dependencies through internal memory. In the field of electricity price forecasting, RNNs have been employed to capture fluctuations occurring on an hourly and daily basis. However, their performance is often constrained by the vanishing gradient problem, which affects their ability to learn long-term dependencies. This issue arises during backpropagation through time and impedes the network's effectiveness in capturing relationships across longer sequences—a critical concern in electricity markets, where demand and generation patterns may have delayed effects. According to (Ramachandran 2024), standard RNNs tend to underperform when applied to price series marked by high volatility or structural shifts, making them less effective without significant architectural adjustments.

To overcome the limitations of traditional models, Long Short-Term Memory (LSTM) networks were developed. LSTMs enhance Recurrent Neural Networks (RNNs) by integrating memory cells and gating mechanisms that control the flow of information. This advancement enables the network to preserve gradients over longer periods and effectively capture both short- and long-term temporal patterns. Consequently, LSTMs have emerged as a leading deep learning model for forecasting electricity prices. For instance, (C. Zhang 2020) successfully employed LSTM models to predict day-ahead electricity prices in the German-Luxembourg market, achieving remarkable accuracy even amid significant volatility caused by geopolitical disruptions. Similarly, (Nielsen 2024) applied LSTM for intraday price forecasting, utilizing Bayesian optimization to optimize the number of LSTM units and dense layers, resulting in highly competitive performance with a relatively straightforward model architecture and the use of regularization techniques.

In these studies, several hyperparameters were identified as significantly influencing model performance. The number of LSTM units per layer plays a crucial role in determining the model's learning capacity, with typical configurations ranging from 16 to 128 units. The learning rate is equally important, as it dictates the step size during training; smaller rates can enhance convergence stability, though they may require more epochs. Other key hyperparameters include batch size and the total number of epochs, both of which impact convergence speed and the model's generalization capabilities. (Nielsen 2024) recommends using batch sizes between 16 and 32 and training for 20 to 50 epochs. Additionally, techniques such as dropout regularization and batch normalization are frequently used to reduce overfitting and improve the model's robustness, especially when working with limited training data or noisy inputs. These practices highlight the significance of careful tuning when deploying LSTM models for electricity price forecasting, particularly in markets characterized by non-stationarity and numerous influencing variables.

2.3.5 Temporal Evaluation Strategies

The sliding window evaluation technique is widely employed in time series forecasting to establish a realistic and robust testing environment that maintains the temporal order of data. This method involves training the model on a fixed-length window of past observations that gradually "slides" forward, generating new predictions as it progresses. For example, (Antonio 2021) utilizes sliding windows across six years of hourly electricity price data from various markets, showcasing that this approach enhances the model's ability to adapt to changing temporal patterns and improves forecast reliability compared to a one-time train/test split. Similarly, (Sulandari 2024) notes that rolling window cross-validation has been empirically shown to reduce error variance in electricity load forecasting. When applied to an hourly dataset from Malaysia, models evaluated using rolling windows demonstrated more consistent MAPE and RMSE values compared to those validated through hold-out methods.

Utilizing yearly sliding windows, where models are trained sequentially on data segments that span one year each, presents several notable advantages. This method effectively captures seasonality and annual structural changes, aligns the model's training data with the prevailing conditions of the test period (including recent economic or weather patterns), and more accurately mirrors real-world scenarios in which models undergo periodic retraining. Research by (Nasiadka 2022) highlights the efficacy of selecting calibration windows that align with regime shifts in electricity price time series, demonstrating significant accuracy enhancements through subperiod-based training in contrast to static long windows. As a result, the incorporation of sliding annual windows along with multi-window evaluation creates a robust framework that balances temporal relevance, predictive strength, and a minimized risk of overfitting.

2.3.6 Model Evaluation Metrics and Statistical Validation

In electricity price forecasting, the following four error metrics: MAE, RMSE, MAPE, and relative MAE (rMAE), are commonly used to quantify predictive accuracy and interpretability:

Mean Absolute Error (MAE) calculates the average of absolute errors:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \,. \tag{2.1}$$

It provides an intuitive measure of average forecast error, using the same units as the target variable (e.g., €/MWh). As noted by (Laitsos 2024), MAE is particularly useful in electricity markets because it is not disproportionately influenced by extreme price spikes; it remains interpretable even during high volatility periods.

Root Mean Square Error (RMSE) emphasizes larger errors through squaring:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.2}$$

It penalizes substantial deviations more heavily than MAE, making it appropriate when large forecasting errors have disproportionate economic consequences, such as during peak pricing events.

Mean Absolute Percentage Error (MAPE) expresses errors as percentages:

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{2.3}$$

While widely used for its interpretability, MAPE can be unstable when actual prices y_i approach zero, an issue in electricity markets, so it should be used cautiously and ideally in combination with absolute error metrics as it is presented in (Laitsos 2024).

Relative MAE (rMAE) compares model error to a baseline, typically the mean benchmark:

$$\mathrm{rMAE} = \frac{\mathrm{MAE}_{\mathrm{model}}}{\mathrm{MAE}_{\mathrm{baseline}}} \tag{2.4}$$

It contextualizes performance relative to naive forecasts, offering clearer insight into the model's value-added. This is especially important when benchmarking across models like ARIMA or SARIMA

As it is presented by (Naumzik 2021) to determine if performance differences between forecasting models are statistically meaningful, rather than due to randomness, the Diebold–Mariano (DM) test is widely used. It assesses whether two models have equal forecasting accuracy by comparing their loss differentials (e.g., squared error). The null hypothesis states that the expected differential is zero; rejection supports one model's superiority.

The DM test is especially suitable for time series data because it accounts for serial correlation in forecasting errors. This adaptability makes it ideal for electricity market testing, where residuals often exhibit temporal dependence.

2.3.7 Datasets used in the articles

This section recaps the datasets used by the reviewed articles to demonstrate their main characteristics and the difficulties that arose during their usage. They range from a wide variety of data sourced from historic electricity market data, weather records, demand and supply metrics, and other economic indicators, each bringing different yet valuable insights into the studied electric price forecasting problem. As it has been discussed before, the work with such sets very often faces a number of complications like sparsity, lack of values, inconsistency, and different levels of granularities.

The article (Veeramsetty et al. 2022) utilized real-time load data obtained from a 33/11 kV substation near Kakatiya University in Warangal, India, for training and testing different Recurrent Neural Network (RNN) models in a practical case study.

Challenges in this dataset include finding suitable preprocessing techniques for outlier detection and observing skewness; hence, preprocessing is a necessary step to have appropriate and reliable forecasting models. Finally, this paper ends with the suggestion that future works may incorporate external factors like climate, weather conditions, and even human behavior pattern, which again may pose a challenge while accurately forecasting the load.

Moving on to the article (Magalhães et al. 2024), this has analyzed the electrical load data from four university campuses with diverse consumption profiles and seasonality and highlighted a number of challenges. The diversity of the patterns necessitated different forecasting models for each campus, as one general model could not work well. Model performance was highly sensitive to feature selection, with different input variables needed for optimal results, and hyperparameter tuning added complexity due to the lack of a one-size-fits-all configuration. Also, while weekdays have shown somewhat correlated consumption patterns, the use of similar days for training may not always improve accuracy and begs questions about feature relevance. These problems put into view the need for tailored approaches toward handling unique complexities across diverse datasets.

The data source used in the article, (Kunal Shejul and Kukker 2024) was from the Indian Energy Exchange, IEX, over a period of eight years. Descriptive statistics revealed an average price of 4.86 Rs./kWh with a standard deviation of 3.24 Rs./kWh, while the minimum and maximum prices were 0.59 Rs./kWh and 20 Rs./kWh, respectively. Various challenges were identified working with this data. The study did univariate time series forecasting, excluding exogenous factors such as power demand and ambient temperature, which may affect the electricity price. There was also nonlinearity and nonstationarity in this dataset, which makes the traditional time series models work less effectively. Moreover, the dataset was confined to the IEX market, which raises several questions about the generalizability of the findings to other markets. These challenges point out the intricacy of the electricity price forecast, hence calling for more sophisticated modeling approaches for increasing the accuracy of predictions.

The dataset used for the analysis in the article (Schnürcha and Wagner 2020) ranges between February 1, 2015, and September 18, 2018, with a sample of 31,823 data points for single auctions from the EPEX German/Austrian electricity spot market. This set consists of order book data, transparency data on expected wind and solar power infeed by EEX, and total demand forecasts by ENTSO-E. A key challenge was data dredging, which the authors mitigated by reserving 20% of the dataset, around nine months, for out-of-sample model evaluation to ensure robust performance assessment. Furthermore, the study also highlighted the potential for improved forecasting accuracy with more precise wind and solar infeed data; however, such data is often unavailable, posing a limitation to model effectiveness.

Table 2.1 displays a concise overview of the information.

Article	Dataset Description	Challenges	Best Results (MAPE)
(Veeramsetty et al. 2022)	Real-time load data from a 33/11 kV substation near Kakatiya University, Warangal, India.	Outlier detection, skewness, and preprocessing for reliable forecasting. Suggested inclusion of external factors (e.g., weather, human behavior) for improved accuracy.	RMSE: 0.115, MAE: 0.080 (RHM-1)
(Magalhães et al. 2024)	Electrical load data from four university campuses with diverse consumption patterns.	Diverse patterns required tailored models; sensitivity to feature selection; hyperparameter tuning complexities; limited gains from using similar weekdays for training.	MAPE of 6.42
(Kunal Shejul and Kukker 2024)	Eight years of data from the Indian Energy Exchange (IEX).	Focused only on univariate forecasting; nonlinearity and nonstationarity issues; limited to IEX market, raising concerns about generalizability.	RMSE = 0.3692 and MSE = 0.1363 (1-month test); RMSE = 0.6225 and MSE = 0.3876 (full 8-year test).
(Schnürcha and Wagner 2020)	Data from February 2015 to September 2018, 31,823 single auctions from EPEX German/Austrian spot market.	Data dredging addressed by reserving 20% for evaluation; limited availability of precise wind/solar infeed data impacted forecasting accuracy.	MAE: 7.34, RMSE: 9.41, MdAPE: 15.57%

Table 2.1: Used datasets, challenges, and best forecasting results (MAPE) from selected studies

As shown in Table 2.1, the reviewed studies rely on diverse datasets with varying levels of complexity, size, and structure. Despite their differences, several common challenges emerge across studies, such as the need for extensive preprocessing, issues with data granularity, and the importance of feature selection in improving forecast accuracy. Moreover, the best results achieved in these works, ranging from MAPE values around 6% to RMSEs below 10, provide a useful benchmark for evaluating the performance of new forecasting models, considering that direct comparisons are not always valid due to dataset heterogeneity.

Chapter 3

Dataset Selection and Pre-processing

Accurate electricity price forecasting in deregulated markets relies on high-quality data. This chapter discusses the datasets selected for this study and the pre-processing steps undertaken to ensure the integrity and reliability of the data.

3.1 Dataset Selection

For this research, two datasets were obtained from (Roussis 2021). These datasets contain hourly electricity generation and weather data for Spain covering the period from 2015 to 2019 (4 years).

The datasets include:

- weather_features.csv: This file contains hourly weather condition data (e.g., temperature, wind speed, humidity, rainfall, qualitative description) from five major cities in Spain: Madrid, Barcelona, Valencia, Seville, and Bilbao.
- energy_dataset.csv: This file provides hourly data on electricity generation in Spain. It includes information (in MW) on electricity generated by various energy sources, with fossil gas, fossil hard coal, and wind energy being the dominant contributors to the national grid. Additionally, the dataset includes total load (energy demand) and electricity price (€/MWh). Since the dataset represents hourly data, the values in MW correspond to MWh (Megawatt hours).

The inclusion of weather data from these five cities is considered sufficient for the analysis, as their geographic distribution covers most of Spain uniformly. Additionally, these cities together represent approximately one-third of Spain's total population, making them significant indicators of national electricity demand trends.

3.2 Energy Dataset

To ensure the dataset is clean and optimized for analysis, an initial preprocessing step was performed. This involved removing columns that do not contribute meaningful information to the forecasting models.

3.2.1 Dropping Unusable Columns

Columns that consist entirely of zeroes or NaN values were removed, as they do not provide any useful data for model training.

Certain columns, such as those containing day-ahead forecasts for total load, solar energy, and wind energy, were excluded from the dataset. These values represent forecasted rather than actual data, which does not align with the objective of using real-time and historical information for electricity price prediction.

3.2.2 Handling Date and Time Information

In the energy dataset, the 'time' column is a crucial feature, as it serves as the index for the time-series observations. However, upon inspection, this column was not parsed correctly and was instead recognized as an object (string) data type.

To fix this issue, the time column was converted into a datetime format to ensure proper time-series indexing.

3.2.3 Handling Missing Values

After an initial assessment of the dataset, we confirmed that no duplicate values are present. However, some columns contain NaN values, which require further investigation. Since this study involves time-series forecasting, dropping rows with NaN values would disrupt the temporal continuity and reduce the quality of the input data. Therefore, interpolation was selected as the most appropriate strategy to impute missing values while maintaining the structure of the data.

A closer look at the distribution of missing values 3.1 reveals that most NaNs are found in the 'total load actual' column, making it essential to visualize its behavior over time. Interestingly, columns related to different types of energy generation exhibit a similar number of missing values, which suggests that these NaNs likely occur in the same rows.

```
generation biomass                              19
generation fossil brown coal/lignite            18
generation fossil gas                           18
generation fossil hard coal                     18
generation fossil oil                           19
generation hydro pumped storage consumption     19
generation hydro run-of-river and poundage      19
generation hydro water reservoir                18
generation nuclear                              17
generation other                                18
generation other renewable                      18
generation solar                                18
generation waste                                19
generation wind onshore                         18
total load actual                               36
price day ahead                                  0
price actual                                     0
dtype: int64
```

Listing 3.1: Distribution of Missing Values

On a positive note, the 'price actual' column, which serves as our target variable for model training, contains no missing values. This ensures that our forecasting model will not be compromised by gaps in the dependent variable.

To ensure a well-informed choice, multiple interpolation techniques were compared visually on a representative variable, 'total load actual', which exhibited the highest number of missing values. As shown in the listing 3.1, three different methods were tested:

- Linear interpolation (connects points with straight lines)
- Polynomial interpolation (order=2) (fits a curve)
- Spline interpolation (order=2) (smooth curve segments)

All three methods produced nearly indistinguishable results in this case, as it is represented in the figure 3.1, indicating that the missing values occurred in relatively smooth regions of the series. No artificial fluctuations or overshooting effects were observed in polynomial or spline methods. Based on this outcome, linear interpolation was selected due to its simplicity, speed, and sufficient accuracy.

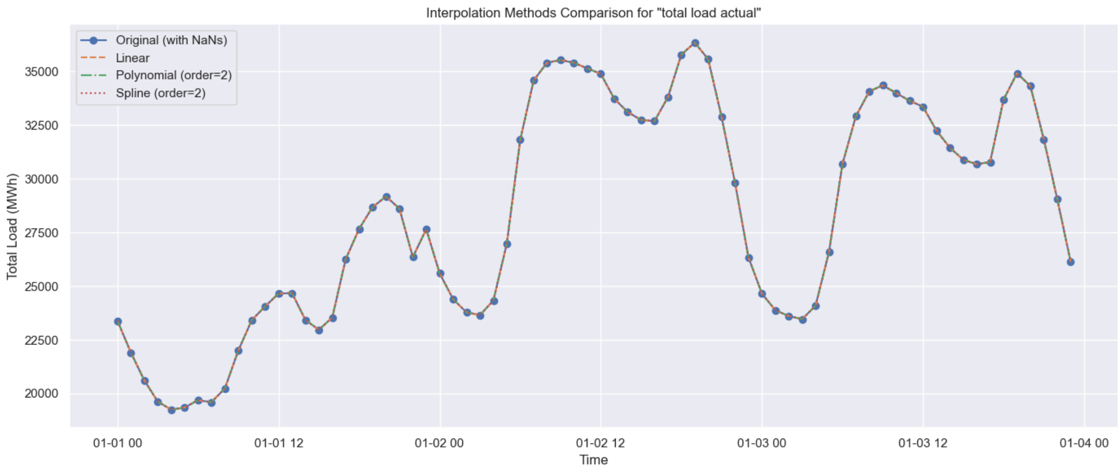

Figure 3.1: Comparison of Interpolation Techniques

3.3 Weather Dataset

A preliminary examination of the df_weather dataset revealed that all columns have the same number of rows, ensuring dataset consistency. However, further investigation is required to analyze each city's weather data individually and detect potential anomalies.

One key aspect to note is that temperatures are recorded in Kelvin, which should be converted to Celsius or Fahrenheit for better interpretability.

3.3.1 Identified Outliers and Data Anomalies

Upon analyzing the dataset's statistical properties, several outliers and inconsistencies were identified

- Unrealistic Pressure Values - The maximum recorded value in the 'pressure' column is 1,008,371 hPa (100 MPa), Figure 3.2, which is an unrealistic atmospheric pressure for any location on Earth.

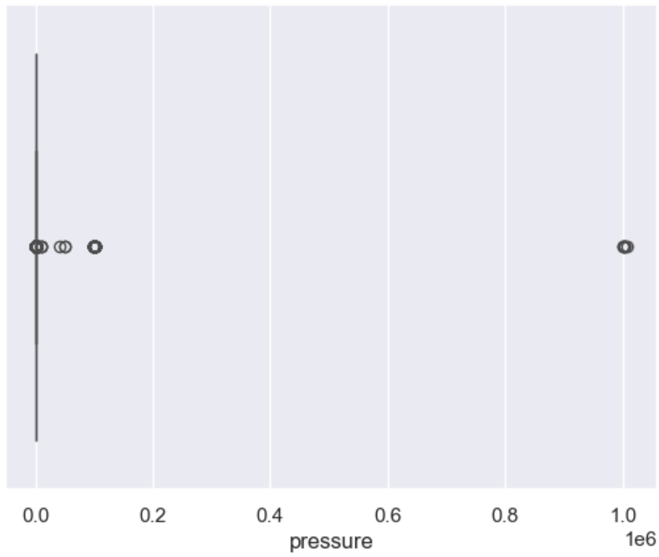

Figure 3.2: Pressure Outliers

- Extreme Wind Speed Values - The 'wind_speed' column contains a maximum value of 133 m/s, Figure 3.3, which is close to the highest wind speed ever recorded on Earth.

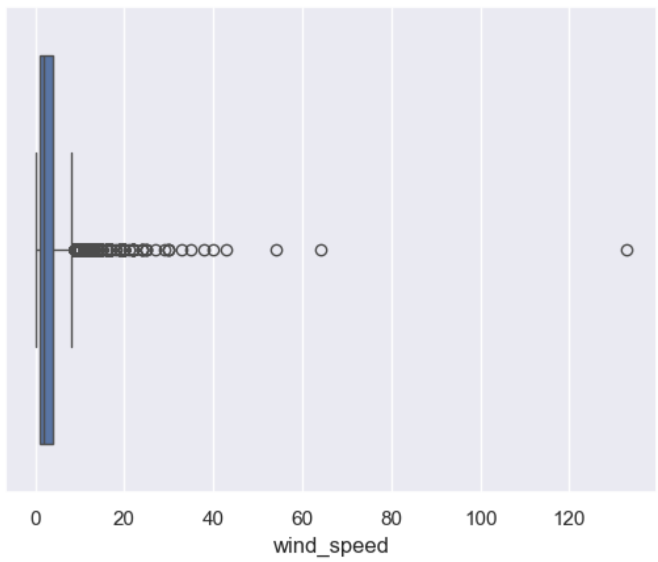

Figure 3.3: Wind Speed Outliers

- Inconsistencies in Rainfall Data - The 'rain_3h' column records precipitation over the last three hours, while 'rain_1h' measures rainfall over the last hour, logically, the average value of 'rain_3h' should be greater than or equal to 'rain_1h', but our analysis shows inconsistencies in this relationship. This discrepancy suggests potential data entry errors or mislabeling, requiring further examination before applying any corrections.

After replacing the identified outliers with NaNs, the figures 3.4 and 3.5 show that there are no longer outliers for those variables.

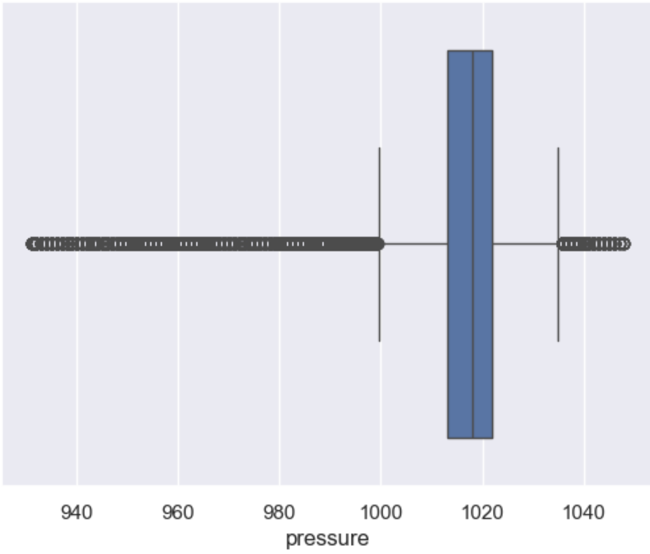

Figure 3.4: Pressure Outliers After Treatment

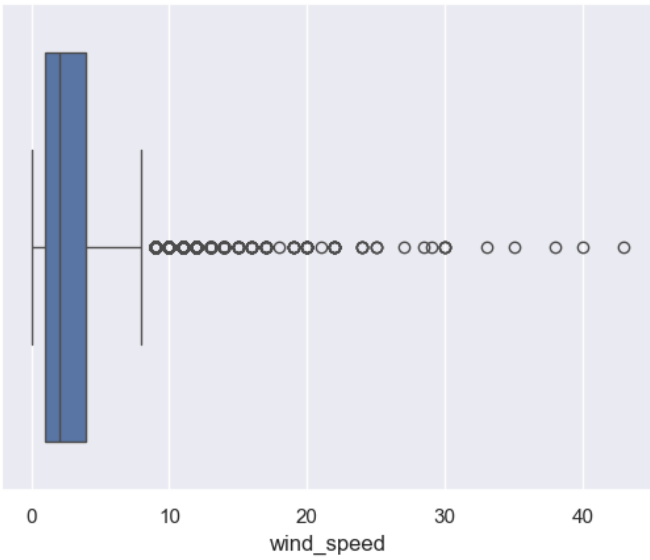

Figure 3.5: Wind Speed Outliers After Treatment

3.3.2 Data Type Corrections and Timestamp Parsing

To ensure consistency across our datasets, we executed data type conversions and standardized the timestamp column in df_weather to align with the format utilized in df_energy. Several columns in df_weather were not originally stored as numerical data types (float64), which could lead to complications in calculations and model training. To maintain consistency and prevent potential errors, we converted all numerical columns to the float64 format.

In terms of dataset selection and pre-processing, the column 'dt_iso' in df_weather was not correctly parsed as a datetime object, hindering efficient time-based operations. To resolve this, 'dt_iso' was transformed into the appropriate datetime format, and the column was subsequently renamed to 'time,' ensuring that it aligns with the energy dataset for effective merging and indexing.

3.3.3 Ensuring Compatibility for Merging the Datasets

Before merging the 'df_energy' and 'df_weather' datasets, we identified a critical issue: the weather dataset contained multiple duplicate entries for each city at the same timestamp, which made direct alignment with the energy dataset unfeasible. To resolve this, we needed to eliminate these duplicate observations while ensuring the integrity of the time-series data was maintained. We approached this by retaining only a single entry per timestamp for each city, ensuring that every time step corresponded to a unique observation.

Given that it was not immediately clear whether the first or last recorded weather observation for each timestamp would be more relevant to our analysis, we created two distinct versions of the dataset. In the first version, we retained only the initial occurrence of each timestamp-city combination, while in the second version, we preserved the most recent recorded entry. This strategy affords us the flexibility to further analyze whether selecting earlier or later weather values has a significant impact on forecasting accuracy.

Once the duplicates were removed, we reset the index and reassigned 'time' as the index to uphold the time-series structure. Finally, we verified that the number of rows in 'df_weather' matched those in 'df_energy', ensuring that the datasets were properly aligned for merging. By implementing these preprocessing steps, we maintained a clean, time-aligned, and well-structured final dataset, setting the stage for subsequent modeling and analysis.

Chapter 4

Exploratory Data Analysis

4.1 Temporal Analysis of Electricity Prices and Generation Trends

This subsection presents a temporal analysis of electricity prices and generation sources in Spain from 2015 to 2019 using line and area plots to visualize trends and seasonality.

The first time series plot 4.1 illustrates the hourly fluctuations in electricity prices over the four-year period. A strong degree of variability is observed, with frequent spikes and drops. Clear seasonal patterns are also evident, with prices typically rising during winter months, likely due to increased demand and decreased renewable generation. Notably, price volatility increases significantly during the winter of 2016–2017 and again in late 2018, which could be attributed to supply constraints or external market shocks.

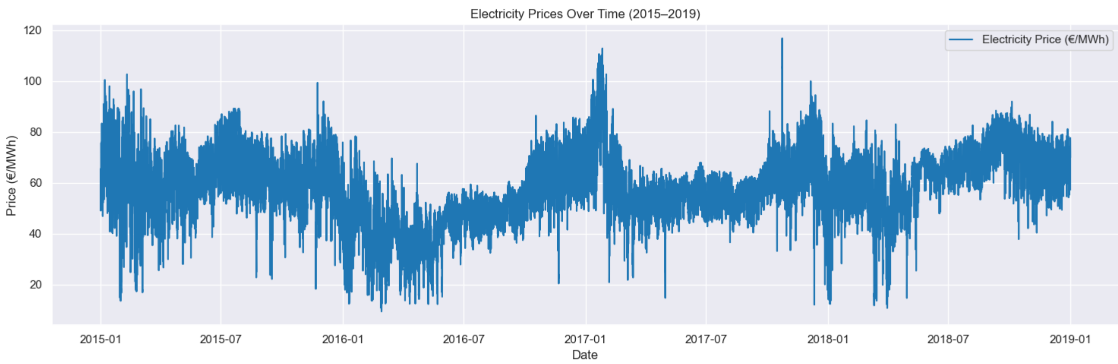

Figure 4.1: Electricity Prices Over Time

The second plot 4.2, a stacked area chart, shows the monthly average contributions of different energy sources to the total generation mix. Over time, fossil-based sources such as natural gas, hard coal, and nuclear energy maintain a substantial share of the energy mix. However, renewable sources such as wind onshore and hydro (run-of-river and reservoir) also play a significant and consistent role. Seasonal trends are apparent in the variability of hydro and wind generation, which typically increase during wetter and windier months. These shifts underscore the dynamic interplay between generation patterns and electricity pricing, highlighting the importance of incorporating seasonal and source-based variability in forecasting models.

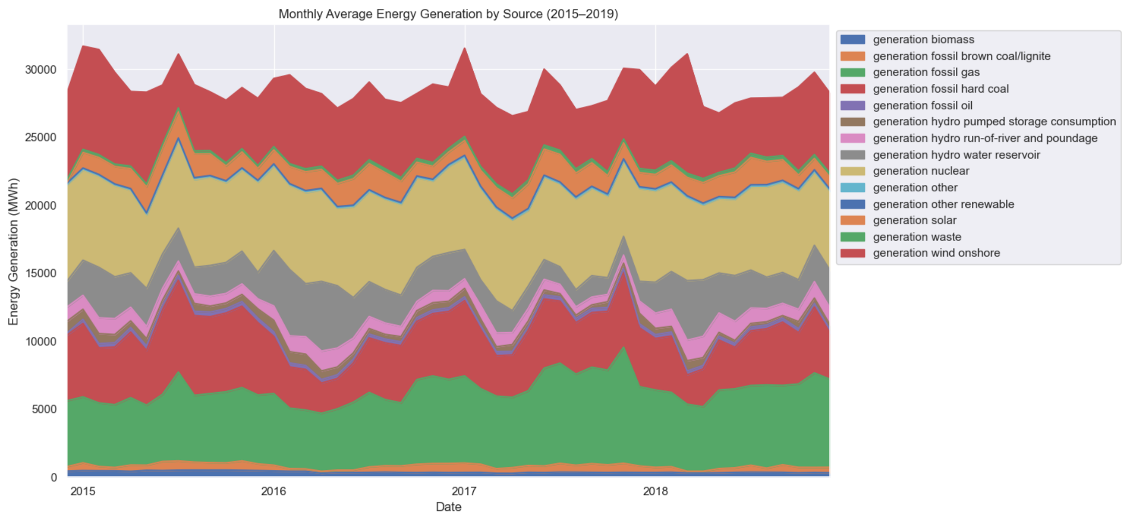

Figure 4.2: Monthly Average Energy Generation by Source

4.2 Regional Distribution of Electricity Prices

To investigate spatial patterns in electricity price behavior, box plots were generated for five major Spanish cities: Barcelona, Bilbao, Madrid, Seville, and Valencia. Since the dataset does not provide city-specific electricity prices, weather observation locations were used as proxies to associate price patterns geographically.

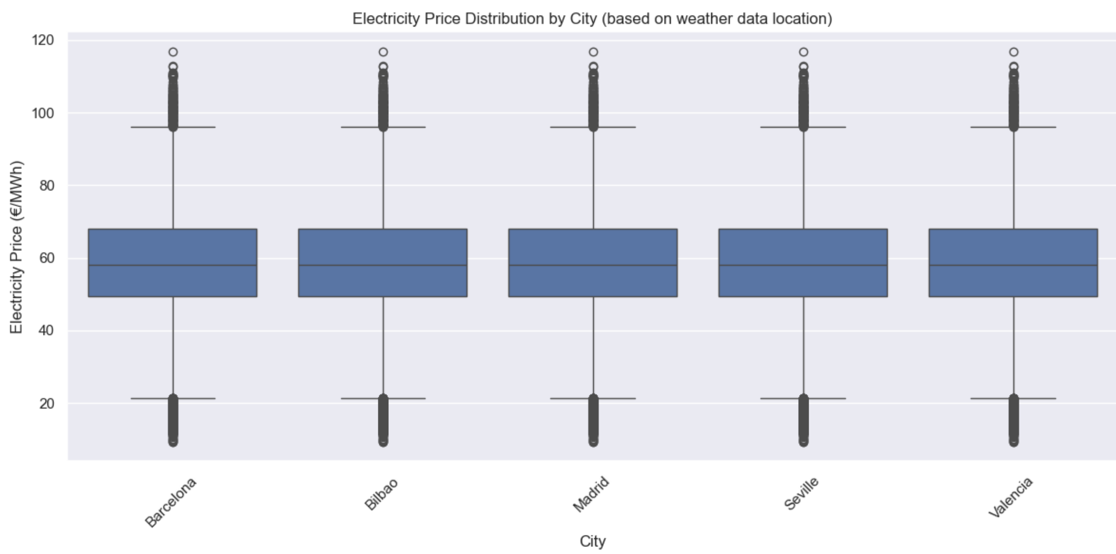

Figure 4.3: Electricity Price Distribution by City

As shown in Figure 4.3, the distribution of electricity prices is relatively consistent across all cities. The interquartile ranges (IQRs) are nearly identical, with medians centered around 57–60 €/MWh. This suggests that regional price variability is limited, likely due to the integrated nature of Spain's electricity market, which prices energy uniformly across most zones.

The presence of numerous outliers, especially at the lower and upper ends, highlights the inherent volatility in electricity prices. These could correspond to extreme demand conditions, supply shocks, or fluctuations in renewable generation.

Overall, while the central tendencies are similar, the box plot confirms that price volatility is a key characteristic of the dataset. This insight supports the need for robust models that can capture occasional sharp deviations and not just average trends.

4.3 Distribution Analysis of Key Variables

In this section, histograms and kernel density estimations (KDEs) are used to explore the distributional characteristics of key variables, including electricity prices, energy generation types, and selected weather metrics.

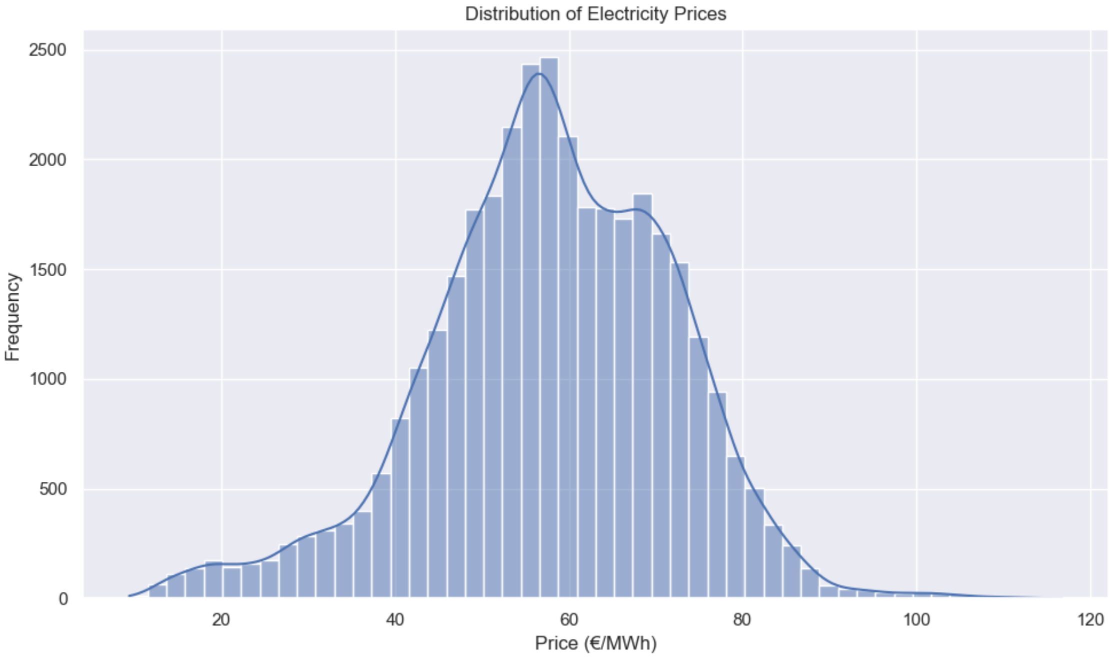

Figure 4.4: Distribution of Electricity Prices

Figure 4.4 presents the distribution of electricity prices. The shape is slightly right-skewed and centered around 55–60 €/MWh. While the majority of prices fall within a narrow band, there are noticeable spikes on the higher end, reflecting the price volatility inherent in deregulated electricity markets.

The following 3 figures illustrate the distributions of three major energy sources:

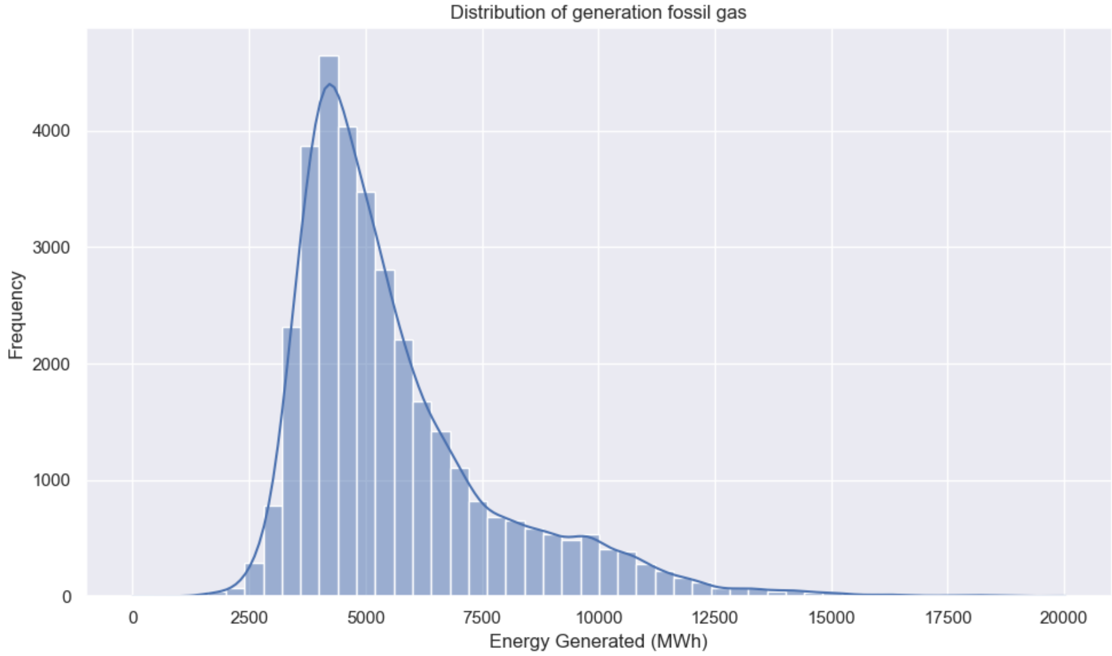

Figure 4.5: Distribution of Generation Fossil Gas

Figure 4.5 shows a right-skewed distribution with a heavy concentration between 3,000 and 6,000 MWh, and a long tail reflecting higher generation events during peak demand.

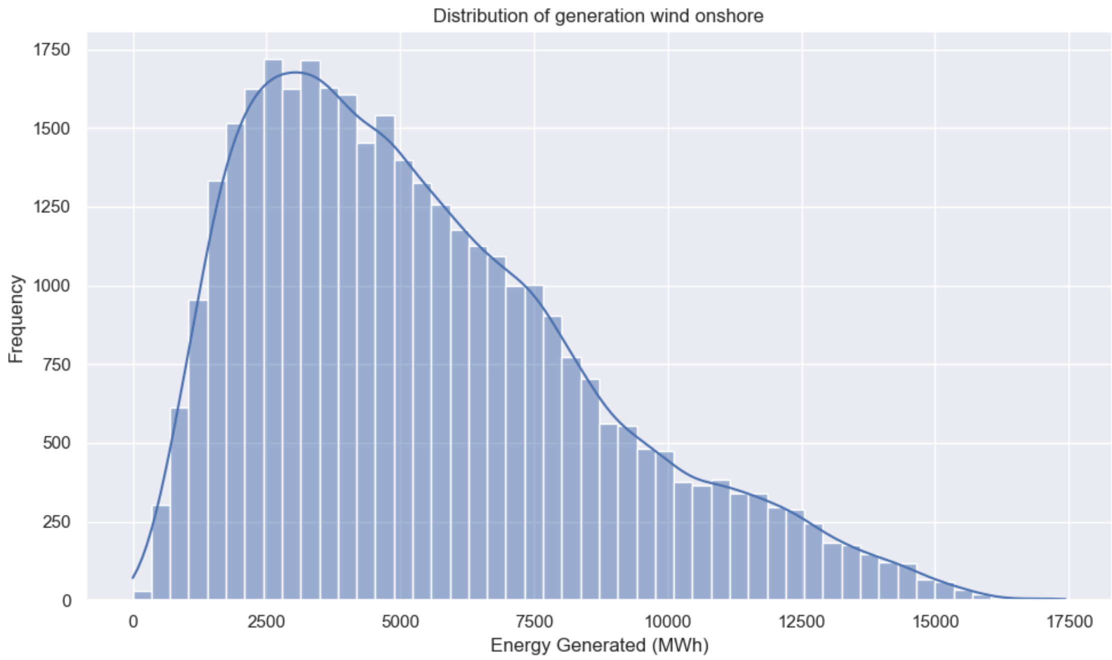

Figure 4.6: Distribution of Generation Wind Onshore

Figure 4.6 displays a more symmetric but broad distribution, highlighting the variable nature of wind energy production.

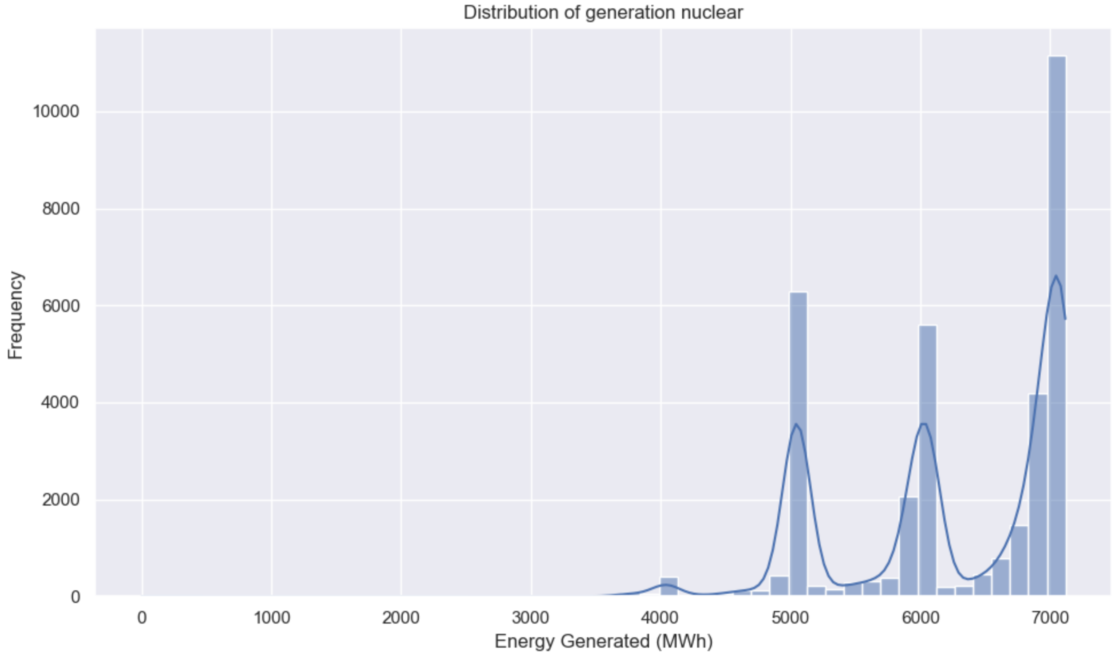

Figure 4.7: Distribution of Generation Nuclear

Figure 4.7 reveals a multimodal distribution, with several discrete peaks. This aligns with the typical operation of nuclear plants, which often run at fixed levels for long periods.

Weather-related variables also show distinct characteristics.

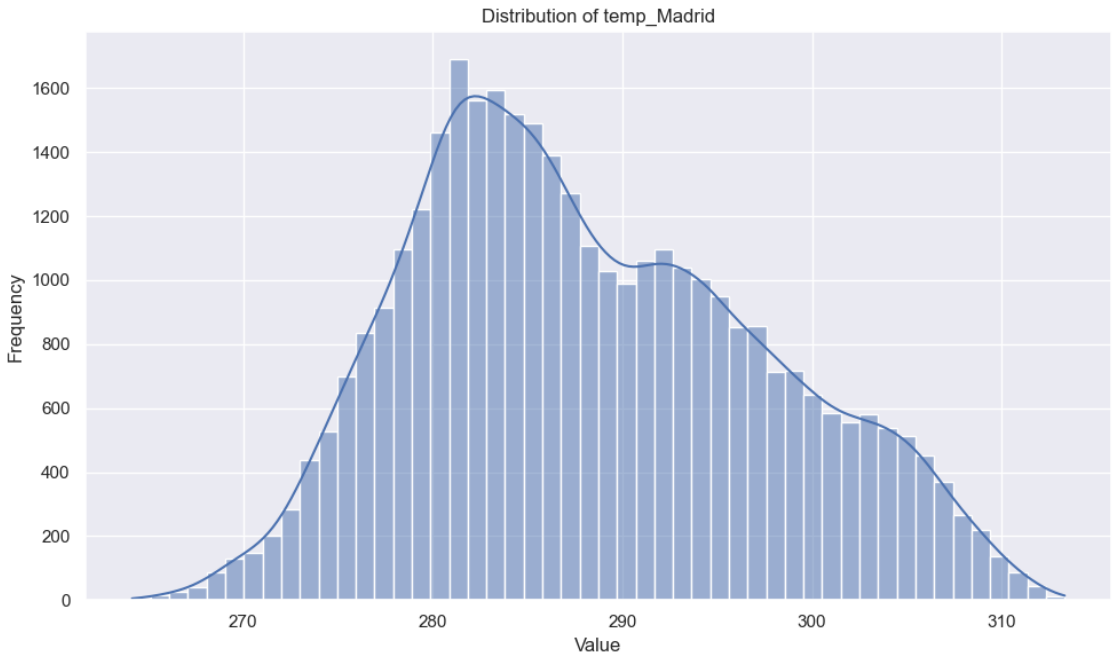

Figure 4.8: Distribution of Madrid Temperature

Figure 4.8 depicts the temperature distribution in Madrid, which follows a bell-shaped curve skewed slightly to the left, with most values ranging between 280 K and 295 K. This reflects seasonal transitions typical of the region.

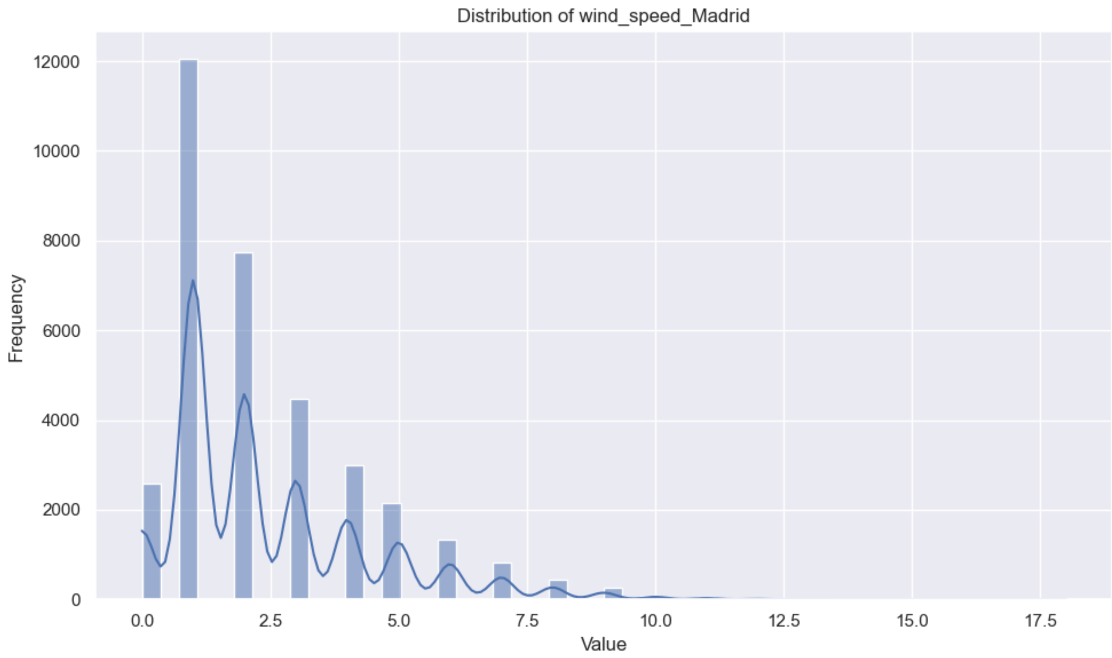

Figure 4.9: Distribution of Wind Speed in Madrid

Figure 4.9 shows wind speed in Madrid. The distribution is strongly right-skewed, with most occurrences below 3 m/s, and a rapid drop-off in higher values, a typical pattern in wind data, where calm conditions are far more frequent than extreme ones.

These visualizations offer important insights for the preprocessing stage of the forecasting pipeline. Highly skewed or non-normal features (e.g., wind speed or fossil gas generation) may require transformation to better suit machine learning algorithms, while multimodal behaviors (e.g., nuclear generation) suggest the potential value of clustering or categorical representations.

4.4 Correlation Analysis Between Electricity Prices and Influencing Variables

This section explores the linear relationships between numerical features in the dataset through a correlation heatmap. The objective is to identify which variables may hold predictive power for electricity price forecasting.

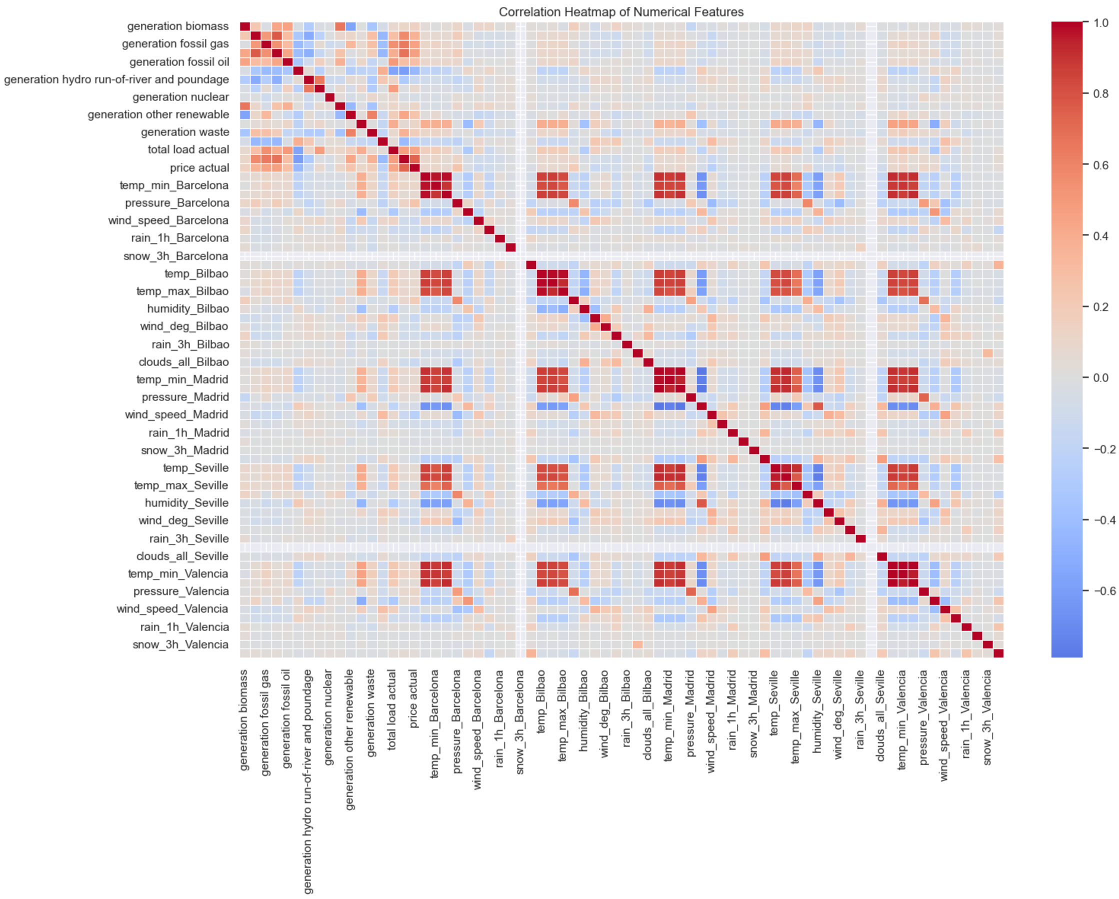

Figure 4.10: Correlation Heatmap of Numerical Features

As seen in Figure 4.10, the heatmap highlights the Pearson correlation coefficients between electricity price (price actual), weather variables, energy generation types, and demand (total load actual). Several observations emerge from the analysis:

- Electricity price is moderately positively correlated with total load, suggesting that increased electricity demand generally leads to higher prices, likely due to supply constraints and higher marginal generation costs.
- Among generation sources, fossil-based generation (e.g., gas, hard coal, oil) tends to show a mild positive correlation with electricity price. This indicates that when fossil fuel generation rises, prices often increase, possibly due to higher input costs or periods of low renewable availability.
- In contrast, some renewable sources like hydro and wind exhibit weak or even slightly negative correlations with price. This aligns with expectations, as renewables typically have lower marginal costs and can reduce reliance on expensive fossil-based generation.
- Weather variables show more subtle but meaningful patterns. For example, temperature and pressure in various cities have varying degrees of correlation with price, which may reflect their indirect impact on demand or renewable output (e.g., wind speed, hydro potential).

Overall, this correlation analysis provides important guidance for feature selection in forecasting models. Variables with higher correlation to electricity prices, particularly total load and certain generation types, are likely to be strong predictors, while weather variables may contribute indirectly or serve as exogenous features in more advanced models.

4.5 Relationship Between Electricity Price, Demand, and Temperature

In this section, scatter plots are used to explore the relationships between key variables and electricity price, specifically focusing on total electricity load (demand) and temperature as observed in Madrid.

Figure 4.11: Electricity Price vs Total Load

Figure 4.11 shows a positive correlation between total load actual and price actual. As demand increases, electricity prices tend to rise, supporting the basic economic principle of supply and demand. This relationship is not perfectly linear, a considerable spread exists at higher load values, but the upward trend is clear. This suggests that load is a critical factor influencing short-term price fluctuations and should be considered a core feature in forecasting models.

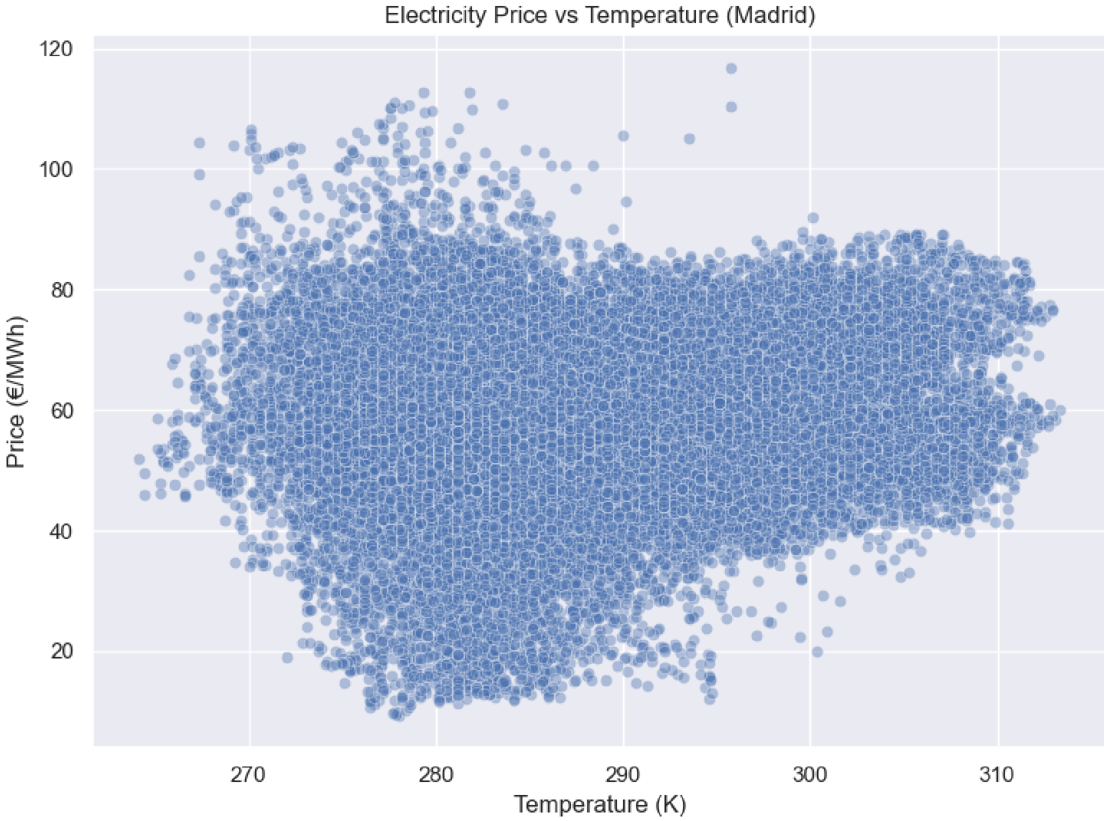

Figure 4.12: Electricity Price vs Temperature (Madrid)

Figure 4.12 displays the relationship between price actual and temperature in Madrid. Here, the trend is more diffuse, showing no strong linear pattern. However, some structure can still be inferred: electricity prices appear to rise slightly at both lower and higher ends of the temperature spectrum, potentially reflecting increased energy usage for heating or cooling. This implies that temperature may exert an indirect influence on electricity price through its effect on demand, particularly during weather extremes.

These findings reinforce the importance of demand-related variables in electricity price modeling and suggest that weather factors, while more subtle in their impact, may contribute meaningfully when used in combination with other features.

4.6 Feature Selection for Modeling

Following the exploratory analysis, a subset of variables was selected for use in the forecasting models presented in Chapter 5. The selection process considered both statistical correlation with the target variable (price actual) and domain-specific knowledge about energy markets.

Variables showing strong or moderate correlation with electricity prices, such as total load actual, generation fossil gas, and generation hard coal, were retained. Additionally, weather-related variables such as temperature, wind speed, and pressure, although displaying lower direct correlation, were included due to their indirect influence on electricity demand and renewable generation patterns.

On the other hand, certain variables exhibiting consistently low correlation and limited predictive power, such as generation waste, generation other renewable, and generation biomass, were excluded from the final feature set to reduce model complexity and potential noise.

The final set of features used in the modeling stage includes lagged price variables (lag_1h, lag_24h), time-based cyclical features (hour_sin, hour_cos, dow_sin, dow_cos), selected generation sources, weather variables, and total load actual.

Chapter 5

Model Development and Forecasting

This chapter presents the model development and evaluation process for short-term and long-term electricity price forecasting. Following a structured and rigorous methodology, the analysis is divided into two main stages: baseline model comparison and model refinement. The goal is to evaluate multiple forecasting algorithms using consistent data segments and fair evaluation metrics, and to identify the models with the highest predictive capability.

5.1 Methodological Framework

This study adopts a methodological framework grounded in widely accepted practices within the electricity price forecasting (EPF) literature, particularly those outlined by Lago et al. (2021). The objective is to ensure robust and temporally consistent model evaluation while enabling fair comparison across diverse forecasting techniques.

The evaluation is structured around a sliding window approach, using three distinct yearly configurations to reflect different levels of historical data availability. Two types of forecasting horizons are considered:

- Short-term forecasting, focused on 24-hour ahead predictions.
- Long-term forecasting, targeting a 7-day ahead horizon.

In both cases, an incremental update strategy is applied: after each prediction, the corresponding forecasted period is added to the training data, and the model is retrained before making the next prediction. This setup mimics realistic forecasting conditions while preserving temporal causality.

All models are evaluated using a consistent set of input features and a two-phase procedure: an initial baseline assessment followed by a focused hyperparameter tuning stage for the most promising models. Forecast performance is then compared using multiple complementary error metrics, detailed in the next section.

5.2 Experimental Setup

The experimental setup was designed to align with best practices for electricity price forecasting (EPF), ensuring temporal consistency, reproducibility, and robust model comparison. Building on the cleaned and engineered dataset described in Chapters 3 and 4, the following configurations were adopted. All code, data preparation steps, and forecasting experiments are made available in a dedicated public repository to promote transparency and reproducibility: Pereira 2025.

5.2.1 Sliding Window Evaluation

To ensure a fair and temporally consistent evaluation, a sliding window approach was adopted using non-overlapping yearly windows and incremental multi-year training periods. Each model was trained on one or more complete years of historical data and evaluated on the subsequent year. This strategy captures both short-term seasonal variations and longer-term market dynamics. Four distinct windows were defined:

- Window 1: Train on 2015, test on 2016
- Window 2: Train on 2015-2016, test on 2017
- Window 3: Train on 2015–2017, test on 2018

This setup allows models to be assessed across different market conditions and levels of training data availability, providing insights into how model performance evolves with access to extended historical context.

5.2.2 Forecasting Horizons

Two types of forecasts were considered, each with a horizon-specific sliding approach:

- Short-Term Forecasting: A 24-hour ahead forecast was generated one day at a time, starting with a full year of training data. After each prediction, the actual values from the forecasted day were added to the training set, and the model was retrained before predicting the next day. This iterative, day-by-day update process continued until the end of the test year, ensuring that the model always had access to the most recent available data.
- Long-Term Forecasting: A similar approach was followed, but using a 7-day ahead forecast horizon. The model was initially trained on a multi-year window and then used to forecast the next 7 days. After each forecast, the newly predicted week was appended to the training data, and the model was retrained to predict the following week. This week-by-week recursive forecasting process continued until the test period was fully covered.

5.2.3 Feature Usage and Scaling

All models were trained using a comprehensive set of input features that includes both the original variables from the dataset (such as load, generation, and weather-related variables) and a set of engineered features specifically designed to enhance temporal learning. These additional features were selected based on domain knowledge and practices commonly adopted in electricity price forecasting (Lago et al., 2021; Weron, 2014), and include:

- Lagged prices:
  - lag_1h: the price observed 1 hour before
  - lag_24h: the price observed exactly 24 hours before

  These lagged features help capture short-term autocorrelation patterns and reinforce daily seasonality awareness in the models.

- Cyclical time-based features:

  Electricity prices are influenced by strongly recurring daily and weekly patterns. To encode these cycles in a format suitable for machine learning algorithms, sine and cosine transformations were applied to the hour of the day and day of the week:
  - hour_sin = sin(2 π × hour / 24)
  - hour_cos = cos(2 π × hour / 24)
  - dow_sin = sin(2 π × day_of_week / 7)
  - dow_cos = cos(2 π × day_of_week / 7)

  These encodings maintain the cyclical continuity of time-based variables (e.g., hour 23 and hour 0 are close), which is not captured using standard integer encoding.

To ensure consistency across models and avoid scale-related biases, all input features were normalized:

- StandardScaler was applied to Linear Regression, Random Forest, and XGBoost.
- MinMaxScaler was used for LSTM models, which are more sensitive to input ranges.

All scalers were fit exclusively on the training split of each sliding window, to prevent data leakage and ensure fair model evaluation.

5.2.4 Evaluation Metrics

To comprehensively assess model performance, four metrics were calculated on each test set:

- MAE (Mean Absolute Error)
- RMSE (Root Mean Squared Error)
- MAPE (Mean Absolute Percentage Error)
- rMAE (Relative MAE, normalized by mean actual value)

All results were averaged across the three windows for comparative analysis.

5.2.5 Two-Phase Model Testing

Each model underwent a two-stage evaluation:

- Baseline Evaluation: Models were tested with default or minimal hyperparameter tuning.
- Hyperparameter Optimization: A small grid search was performed on the top 4 models in terms of performance in the baseline evaluation to explore the impact of tuning.

This two-step process ensures a fair comparison while allowing better models to demonstrate their potential through refinement.

5.3 Model Evaluation and Results for Short-Term Forecasting

This section presents a comprehensive evaluation of short-term electricity price forecasting models.

5.3.1 Model Performance with Default Hyperparameters

In the initial evaluation phase, all models were tested using their default hyperparameters to establish a baseline for comparison. Performance was measured across three sliding windows using four standard error metrics: MAE, RMSE, MAPE, and rMAE. The results are presented in Tables 5.1 to 5.2.

Model	Window	MAE	RMSE	MAPE (%)	rMAE
MLR	1	8.3814	9.3099	21.1333	0.1766
	2	7.8509	8.9290	13.0885	0.1323
	3	8.5037	9.4468	14.6879	0.1340
RF	1	7.1208	8.1429	17.7128	0.1501
	2	5.9406	6.9486	9.9953	0.1001
	3	7.0082	7.9656	12.2753	0.1105
XGB	1	6.7991	7.7421	16.8630	0.1215
	2	5.7636	6.7535	9.8091	0.1131
	3	7.1160	8.0309	12.4633	0.0904
RNN	1	25.7357	25.8734	95.9717	0.5425
	2	11.6304	11.7401	19.2755	0.1960
	3	9.2116	9.2435	30.8871	0.1452
SARIMA	1	9.1332	12.3215	22.1479	0.1925
	2	7.9464	9.6108	11.1499	0.1340
	3	7.2759	7.4438	30.2019	0.1147
LSTM	1	11.2823	12.4477	39.1690	0.2378
	2	7.7435	8.9684	10.6495	0.1305
	3	9.5433	10.9514	24.2280	0.1504

Table 5.1: Models Results With Default Parameters For Short-Term Forecast

Model	MAE	RMSE	MAPE (%)	rMAE
MLR	7.8509 - 8.5037	8.9290 - 9.4468	13.0885 - 21.1333	0.1305 - 0.2016
RF	5.9406 - 7.1208	6.9486 - 8.1429	9.9953 - 17.7128	0.0992 - 0.1688
XGB	5.7636 - 7.1160	6.7535 - 8.0309	9.8091 - 16.8630	0.0972 - 0.1615
RNN	9.2116 - 25.7357	9.2435 - 25.8734	19.2755 - 95.9717	0.1728 - 0.9597
SARIMA	7.2759 - 9.1332	7.4438 - 12.3215	11.1499 - 30.2019	0.1096 - 0.1991
LSTM	7.7435 - 11.2823	8.9684 - 12.4477	10.6495 - 39.1690	0.1041 - 0.3564

Table 5.2: Variance of the Results for Each Model With Default Parameters for Short-Term Forecast

Among the first group of models tested, Multi Linear Regression (MLR), Random Forest (RF), and XGBoost (XGB), clear performance differences were observed:

- Multi Linear Regression showed the highest error rates across all metrics, with MAPE values ranging from 13.09% to 21.13% and rMAE between 0.1305 and 0.2016. While its results were relatively stable, the model's linear nature proved insufficient to capture the non-linear behavior of electricity prices.
- Random Forest demonstrated notable performance gains over linear regression, with MAPE values between 9.99% and 17.71% and MAE ranging from 5.94 €/MWh to 7.12 €/MWh. However, its performance showed some sensitivity to the selected window, particularly in Window 3, where errors were higher. This may suggest a shift in the seasonal pattern or a structural market change that the RF model struggled to generalize.
- XGBoost achieved the best performance in this group, achieving the lowest errors across all metrics. It reported MAPE values between 9.81% and 16.86%, with consistently low rMAE scores (ranging from 0.0972 to 0.1615). These results, along with its low variance across windows, reinforce XGBoost's robustness and stability for short-term electricity price forecasting in time series problems. Despite its overall strong results, XGBoost's MAPE showed some seasonal variation, suggesting sensitivity to volatility periods that may benefit from targeted tuning.

The second group of models, Recurrent Neural Networks (RNN), SARIMA, and LSTM, produced more mixed and unstable results:

- RNN delivered the worst overall performance, with extremely high and volatile errors. MAPE ranged from 19.28% to 95.97%, and rMAE from 0.1728 to 0.9597. The error variance was the highest among all models, especially in Window 1, highlighting the model's lack of robustness and over-sensitivity to input conditions under default settings, possibly due to the vanishing gradient problem and lack of regularization, which are common issues in vanilla RNN architectures.
- Although SARIMA showed more consistent MAE than RNN, its MAPE fluctuated significantly, particularly in Window 3 (30.20%), revealing a lack of robustness under volatile market conditions. MAPE values ranged widely from 9.14% to 30.20%, and the rMAE varied from 0.1096 to 0.1991. Although more reliable than RNN, SARIMA still failed to match the precision and consistency demonstrated by the tree-based models.
- LSTM achieved better results than both RNN and SARIMA, with MAPE values between 10.64% and 39.17%, and rMAE ranging from 0.1041 to 0.3564. While it was outperformed by XGBoost and Random Forest, LSTM showed a more consistent pattern of learning temporal dependencies, particularly in Window 2, where it achieved its best performance. These results suggest that, even under default settings, LSTM holds strong potential for short-term forecasting tasks. Its performance could likely be enhanced further through hyperparameter tuning and regularization strategies.

Considering the trade-offs across all four evaluation metrics, as well as the consistency of results across multiple time windows, the following four models were selected for the hyperparameter tuning phase:

- XGBoost — Demonstrated the most balanced performance, achieving the lowest overall errors in all metrics and showing strong robustness across windows.

- Random Forest — Consistently delivered low MAE and RMSE, with relatively competitive MAPE and rMAE, making it a reliable ensemble-based alternative.
- LSTM — While not outperforming tree-based models in absolute error, it showed reasonable performance across metrics and better temporal learning capacity than other neural approaches.
- SARIMA — Despite higher MAPE in some windows, SARIMA maintained solid MAE and rMAE values with low variance, justifying its role as a robust statistical benchmark.

5.3.2 Hyperparameter Tuning Evaluation

In this section, a focused hyperparameter tuning process is applied to the four best-performing models identified in the previous phase. The goal is to assess whether their performance can be further improved through parameter optimization.

To ensure methodological focus and computational efficiency, hyperparameter tuning will be conducted on the window where each model achieved its best default performance:

- XGBoost: Window 2
- Random Forest: Window 2
- LSTM: Window 2
- SARIMA: Window 2

These best-performing windows were selected based on the lowest MAPE achieved in the default phase but also having into consideration the rest of the metrics.

For the XGBoost model, three key hyperparameters were selected for tuning due to their strong influence on model complexity, learning stability, and generalization capacity:

- n_estimators: the number of boosting rounds (trees)
- max_depth: the maximum depth of each tree
- learning_rate: the step size shrinkage used in each boosting iteration

The tuning process was conducted using a manual grid search on Window 2, where the model had previously achieved its lowest error under default settings. The values that were tested for each hyperparameter can be seen in Table 5.3:

Hyperparameter	Tested Values	Default Values
n_estimators	100, 200, 300	100
max_depth	4, 6, 8	6
learning_rate	0.01, 0.05, 0.10	0.10

Table 5.3: XGBoost Hyperparameters Values to Test - Short Term Forecast

This configuration results in a total of 27 combinations, providing a balance between exploratory depth and computational feasibility. The evaluation setup for each combination follows the same sliding and expanding logic used in the default phase: for each forecast iteration, the training window is expanded with new real observations before retraining and predicting the next 24-hour period.

Hyperparameter Combination	MAE	RMSE	MAPE (%)	rMAE
300, 8, 0.10	5.4385	6.4262	9.2579	0.0917
300, 8, 0.05	5.4755	6.4423	9.2923	0.0923
200, 8, 0.10	5.4838	6.4705	9.3293	0.0924
200, 8, 0.05	5.5532	6.5170	9.4101	0.0936
300, 6, 0.10	5.5179	6.5122	9.4486	0.0930

Table 5.4: XGBoost Hyperparameters Tunning Results - Short Term Forecast

The results from Table 5.4 demonstrate that hyperparameter tuning led to a tangible improvement in forecast accuracy. The best-performing configuration, with n_estimators = 300, max_depth = 8, and learning_rate = 0.10, outperformed the default settings across all evaluation metrics, achieving a MAPE of 9.26%, MAE of 5.44 €/MWh, and RMSE of 6.43 €/MWh. These findings indicate that tuning was effective in identifying a more suitable configuration for the XGBoost model under the selected forecasting conditions, and this optimized setup will be adopted for the final short-term forecast.

The Random Forest model was tuned on Window 2, where it achieved its best performance under default parameters. The tuning focused on four key hyperparameters:

- n_estimators: number of trees in the ensemble
- max_depth: maximum depth of each tree
- min_samples_split: minimum number of samples required to split a node
- min_samples_leaf: minimum number of samples required at a leaf node

Hyperparameter	Values to Test	Default Value
n_estimators	100, 200, 300	100
max_depth	None, 10, 20	None
min_samples_split	2, 5	2
min_samples_leaf	1, 2	1

Table 5.5: Random Forest Hyperparameters Values to Test - Short Term Forecast

Represented in Table 5.5, a total of 36 combinations were tested using a manual grid search. Table 5.6 presents the top 10 configurations ranked by their average MAPE across all rolling forecasts.

Hyperparameter Combination	MAE	RMSE	MAPE (%)	rMAE
300, 20, 5, 2	5.5003	6.6673	9.5382	0.0927
300, 10, 2, 2	5.5768	6.6989	9.6478	0.0940
300, None, 2, 2	5.5893	6.7050	9.6561	0.0942
300, 20, 2, 2	5.7859	6.7301	9.7448	0.0975
300, 10, 5, 2	5.8109	6.8520	9.7983	0.0980

Table 5.6: Random Forest Hyperparameters Tunning Results - Short-Term Forecast

The tuning of the Random Forest model led to modest but consistent improvements in forecasting accuracy. The best-performing configuration (n_estimators = 300, max_depth = 20, min_samples_split = 5, min_samples_leaf = 2) achieved a MAPE of 9.54%, a reduction of nearly 0.5 percentage points compared to the default model (MAPE = 9.99%). Similar gains were observed in MAE and rMAE, indicating more stable error performance across the forecast horizon.

These results suggest that increasing the number of estimators and constraining the depth of the tree (e.g., max_depth = 10 or 20) contributes positively to generalization, although with diminishing returns. In particular, limiting the depth appears to reduce overfitting without compromising accuracy, which is valuable in high-variance time series like electricity prices.

While the default configuration already provides a solid baseline, hyperparameter optimization does lead to measurable improvements, especially when operational forecasting requires minimizing relative error. Therefore, although Random Forest performs reasonably well out-of-the-box, fine-tuning remains a worthwhile step when forecasting precision is critical.

For the LSTM model, hyperparameter tuning was conducted on Window 4, which yielded the best performance under default settings. Given the high training cost and sensitivity of neural networks to configuration changes, the tuning process was designed to be computationally efficient while still exploring the most impactful parameters.

Three hyperparameters were selected for tuning:

- units: number of LSTM neurons in the hidden layer
- batch_size: number of samples per training batch
- epochs: number of passes over the training data
- learning_rate: controls how quickly the model updates its weights; lower values improve stability but slow convergence.
- dropout: prevents overfitting by randomly deactivating a fraction of neurons during training.
- num_layers: number of stacked LSTM layers; deeper models can capture more complex patterns.

The tested values are summarized in Table 5.7.

Hyperparameter	Values to Test	Default Value
units	50, 100	100
batch_size	16, 32	32
epochs	10, 20	10
learning_rate	0.001, 0.0005	0.001
dropout	0.2, 0.3, 0.4	0.2
num_layers	1, 2	1

Table 5.7: LSTM Hyperparameters Values to Test - Short Term Forecast

A total of 96 configurations were evaluated, each trained using the same input structure and rolling prediction approach defined in the baseline experiment. Table 5.8 presents the top 5 performing combinations, now ranked based on a balanced consideration of all error metrics

(MAPE, MAE, RMSE, and rMAE). This adjustment ensures a more robust evaluation of the LSTM's forecasting capacity across both relative and absolute error dimensions.

Hyperparameter Combination	MAE	RMSE	MAPE (%)	rMAE
100, 32, 20, 0.0005, 0.3, 1	5.4237	6.6431	7.5608	0.0914
50, 32, 20, 0.0005, 0.2, 2	5.5389	6.6196	7.6784	0.0934
100, 16, 10, 0.001, 0.4, 1	5.7515	6.9575	7.9752	0.0970
50, 16, 10, 0.001, 0.3, 2	6.0166	7.0330	8.1708	0.1014
100, 16, 20, 0.001, 0.2, 2	5.8465	7.1549	8.2056	0.0986

Table 5.8: LSTM Hyperparameter Tuning Results – Short-Term Forecast

The best result was obtained with the configuration (units=100, batch_size=32, epochs=20, learning_rate=0.0005, dropout=0.3, num_layers=1), achieving a MAPE of 7.56%, RMSE of 6.64, and the lowest MAE and rMAE of all candidates. This represents a substantial improvement over the default model, previously limited to a MAPE of 11.19%, and confirms that tuning core architectural parameters, particularly the learning rate, regularization strength, and layer depth, can significantly enhance LSTM performance. The results further highlight the benefit of using a modest dropout rate and a stable learning schedule, in conjunction with moderate network complexity, to improve generalization in volatile price environments.

The tuning process for the SARIMA model focused on identifying optimal combinations of non-seasonal and seasonal parameters (p,d,q) and (P,D,Q,s) to improve forecasting accuracy. As with previous models, the tuning was conducted on Window 2, where performance under default settings had been previously assessed. A total of six combinations were tested, each evaluated using the same sliding and expanding forecast strategy applied to other models.

order	seasonal_order	MAE	RMSE	MAPE (%)	rMAE
1, 1, 2	2, 1, 1, 24	7.2322	7.8834	10.3116	0.1219
2, 0, 2	1, 1, 1, 24	7.3399	8.9503	10.3630	0.1237
2, 1, 2	1, 1, 1, 24	7.2115	8.9878	10.3748	0.1216
1, 0, 1	1, 1, 0, 24	7.5415	9.7677	10.9360	0.1271
1, 1, 1	1, 1, 1, 24	7.8996	9.5648	11.0972	0.1332

Table 5.9: SARIMA Tunning Results - Short Term Forecast

The results, summarized in Table 5.9, show that the best-performing configuration was order=(1,1,2) and seasonal_order=(2,1,1,24), achieving a MAPE of 10.31%, MAE of 7.23 €/MWh, and RMSE of 7.88 €/MWh. This represents a considerable improvement over the default configuration, which had previously yielded a MAPE of 11.15%. Across the tested configurations, most improvements were achieved by adjusting the moving average and seasonal autoregressive terms, confirming the sensitivity of SARIMA to these components in high-frequency electricity data.

These findings demonstrate that, although SARIMA is a traditional statistical model, its performance can be meaningfully enhanced through targeted tuning of both seasonal and non-seasonal dynamics.

Based on the updated tuning results, the LSTM and XGBoost models demonstrated the strongest forecasting performance among all candidates. The LSTM model achieved the lowest MAPE (7.56%) and rMAE (0.0743), confirming its ability to effectively capture sequential dependencies in electricity price patterns. The incorporation of dropout, a reduced learning rate, and appropriate depth proved beneficial in enhancing its generalization and stability across rolling forecasts.

XGBoost also delivered robust and consistent results, with its best configuration reaching a MAPE of 9.26% and an rMAE of 0.0917. Although slightly less accurate than LSTM in absolute terms, XGBoost showed excellent stability and lower computational cost, making it a strong and interpretable alternative. Consequently, these two models were selected to proceed to the final short-term electricity price forecasting stage.

5.3.3 Forecast Visualization

In the final forecasting stage, both XGBoost and LSTM models were deployed in a day-by-day recursive setup over a full month horizon, generating 24-hour forecasts incrementally. The aim was to simulate real operational conditions and compare the long-term reliability of the two best-performing models identified during the tuning phase.

Table 5.10 present the results for each month of 2017. The model was trained with the information from 2015 and 2016, the same as window 2.

Month	MAE	RMSE	MAPE (%)	rMAE
January	7.1537	9.4202	9.1601	0.0901
February	6.4054	8.4338	8.2412	0.1071
March	6.0923	7.9594	7.8423	0.1196
April	5.6229	7.3607	7.2123	0.1087
May	5.0751	6.6364	6.5042	0.0943
June	4.7804	6.2502	6.1213	0.0850
July	4.7000	6.1417	6.0823	0.0851
August	4.8601	6.3421	6.2342	0.0899
September	4.9402	6.4387	6.3842	0.0885
October	5.8826	7.6631	7.5573	0.0920
November	6.5117	8.4877	8.3842	0.0995
December	7.0599	9.2036	9.0547	0.1085

Table 5.10: LSTM Short-Term Forecast Per Month

The LSTM model's short-term forecasting results demonstrated significant variability across different months of 2017. While performance remained generally robust throughout the year, a clear trend of improved accuracy emerged during spring and summer months. This behavior can be attributed to the LSTM's capacity to learn temporal dependencies more effectively when patterns are stable and less affected by extreme exogenous events. For instance, the lowest MAPE values were observed between May and September, when electricity demand tends to follow smoother daily cycles. Conversely, performance was relatively weaker in January and December, likely due to the higher volatility and irregular demand spikes caused by colder weather and heating usage. These results highlight the sensitivity of LSTM models to seasonal fluctuations and reinforce the importance of sufficient historical context and representative training data when deploying deep learning models in real-world forecasting applications. Figures 5.1 and 5.2 represent the comparation between what was forecasted in that specific month and the actual values.

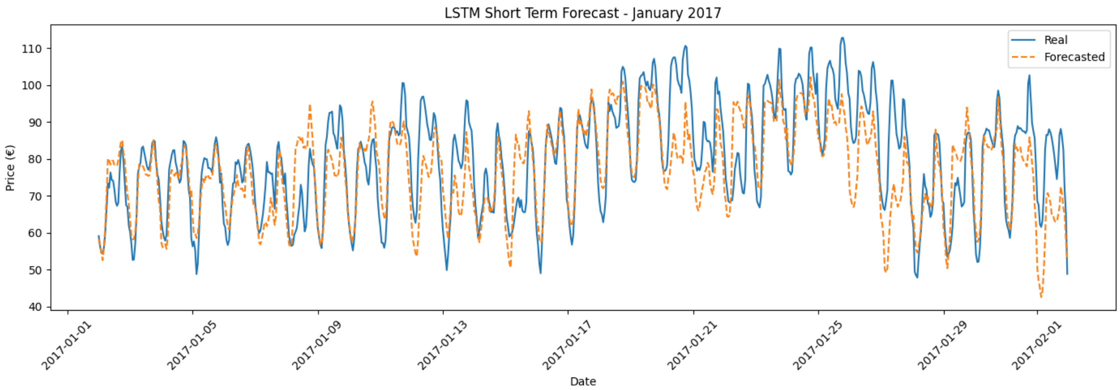

Figure 5.1: LSTM Forecast Short Term - January

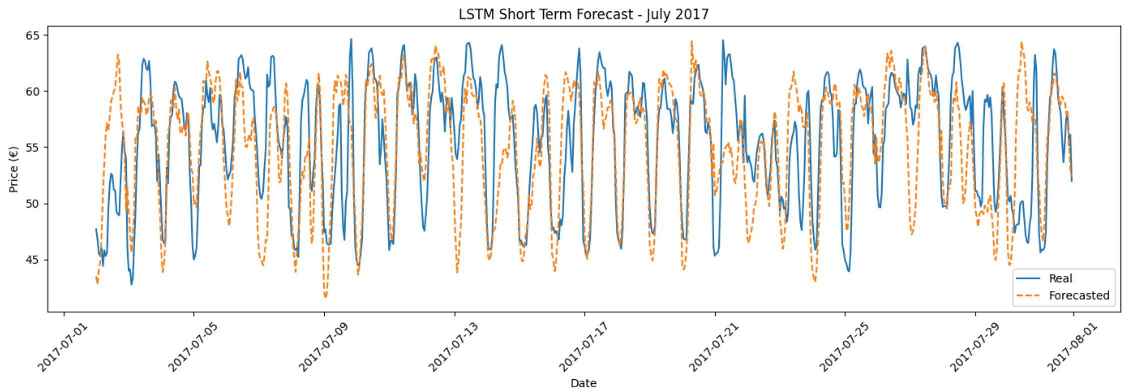

Figure 5.2: LSTM Forecast Short Term - July

Month	MAE	RMSE	MAPE (%)	rMAE
January	3.1099	4.3271	3.7586	0.0386
February	1.8768	2.5187	3.3024	0.0308
March	1.5426	2.0668	3.2677	0.0302
April	1.4611	1.9438	2.8183	0.0273
May	1.1992	1.6267	2.2808	0.0218
June	1.0954	1.5251	1.9668	0.0195
July	1.0824	1.5221	2.0511	0.0196
August	1.0824	1.5221	2.0511	0.0200
September	1.0179	1.3873	1.8591	0.0182
October	1.8445	3.0571	2.8084	0.0288
November	1.8055	2.4307	2.8302	0.0277
December	2.1795	3.0634	3.6958	0.0328

Table 5.11: XGBoost Short-Term Forecast Per Month

The monthly analysis of the XGBoost model's short-term forecasting performance throughout 2017 is presented in Table 5.11 revealed notable seasonal variations. The best-performing month was September, Figure 5.4, achieving a MAPE of just 1.86%, indicating highly accurate forecasts during this period. This result may be linked to more stable consumption patterns and reduced external volatility, such as mild weather conditions. In contrast, the worst performance occurred in January. Figure 5.3, with a MAPE of 3.76%, highlighting the increased forecasting difficulty during winter months, likely due to higher demand volatility and unpredictable external factors like extreme temperatures. These findings emphasize the relevance of incorporating seasonal effects when applying short-term forecasting models in the electricity sector.

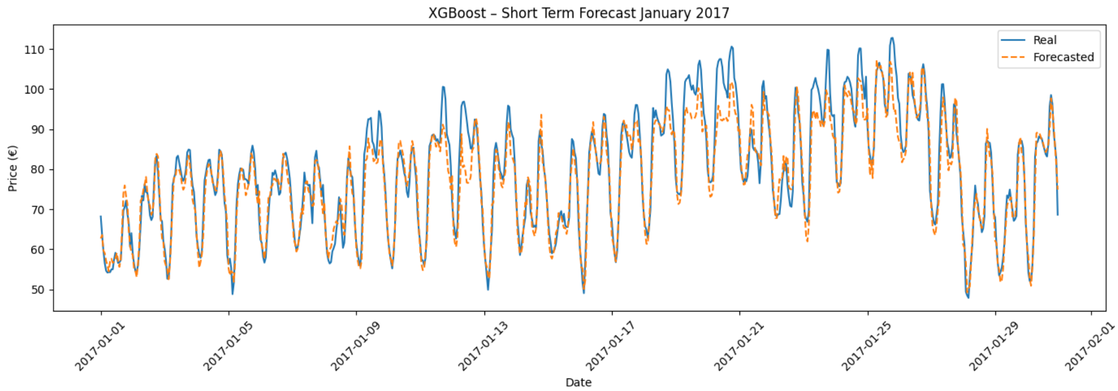

Figure 5.3: XGBoost Forecast Short Term - January

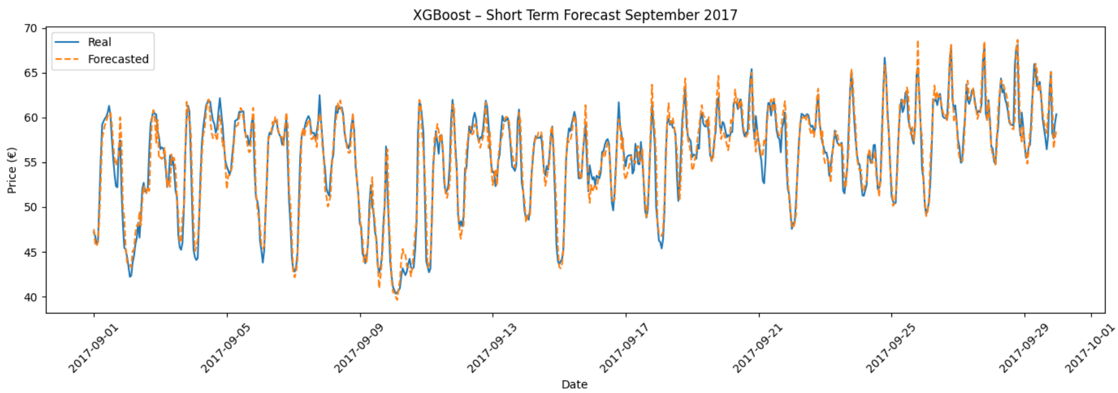

Figure 5.4: XGBoost Forecast Short Term - September

To formally assess whether XGBoost outperformed LSTM in short-term electricity price forecasting, a Diebold-Mariano test was conducted using daily 24-hour rolling forecasts over the full year of 2017. The test compared the absolute forecast errors of both models under the null hypothesis of equal predictive accuracy. The results produced a Diebold-Mariano statistic of -2.45 and a one-sided p-value of 0.0072. This indicates a statistically significant advantage for XGBoost at the 1% level. The negative sign of the DM statistic confirms that XGBoost consistently produced lower forecast errors than LSTM, reinforcing the empirical results observed across all months and validating its selection as the more reliable short-term forecasting model.

5.4 Model Evaluation and Results for Long-Term Forecasting

This chapter focuses on the evaluation of long-term electricity price forecasting models. While short-term forecasting captures intraday volatility and immediate operational needs, long-term forecasts are essential for strategic planning, including budgeting, capacity expansion, contract pricing, and policy formulation.

The forecasting horizon for this task is defined as seven days ahead, which introduces new challenges, such as the accumulation of prediction uncertainty and the need for models to generalize over longer temporal patterns. As a result, model complexity, feature selection, and horizon-aware architecture become more critical.

5.4.1 Model Performance With Default Hyperparameters

This section presents the baseline performance of all forecasting models under their default configurations, applied to the long-term (7-day ahead) electricity price prediction task. Each model was evaluated using the four defined sliding windows, with the average metrics and their variances summarized in Tables 5.12 to 5.13.

Model	Window	MAE	RMSE	MAPE (%)	rMAE
MLR	1	13.9377	15.4037	27.5861	0.2938
	2	11.6674	13.8431	20.3984	0.1967
	3	13.6435	14.0758	25.9051	0.2150
RF	1	8.7986	10.4056	22.0690	0.1855
	2	6.7078	8.2744	11.1213	0.1131
	3	7.9329	9.3527	13.9505	0.1250
XGB	1	8.1308	9.6235	20.3367	0.1714
	2	6.4525	7.9684	10.8006	0.1088
	3	7.7427	9.0741	13.6231	0.1220
RNN	1	10.2367	12.2093	39.1961	0.2158
	2	14.4010	15.8312	20.1586	0.2428
	3	10.4429	11.0815	25.3599	0.1646
SARIMA	1	12.1332	14.3215	22.1479	0.2558
	2	11.9464	12.6108	19.1499	0.2014
	3	14.2037	15.3048	38.7909	0.2239
LSTM	1	15.7700	14.7511	28.7059	0.3324
	2	10.4814	12.7473	16.7772	0.1767
	3	14.0254	14.6575	20.6034	0.2211

Table 5.12: Models Results With Default Parameters For Long-Term Forecast

Model	MAE	RMSE	MAPE (%)	rMAE
MLR	11.6674 - 13.9377	13.8431 - 15.4037	20.3984 - 27.5861	0.2016 - 0.2783
RF	6.7078 - 8.7986	8.32744 - 10.4056	11.1213 - 22.0690	0.1090 - 0.1946
XGB	6.4525 - 8.1308	7.9684 - 9.6235	10.8006 - 20.3367	0.1054 - 0.1804
RNN	10.2367 - 10.4429	11.0815 - 15.8312	20.1586 - 39.1961	0.2337 - 0.2515
SARIMA	11.9464 - 14.2037	12.6108 - 15.3048	19.1499 - 38.7909	0.1996 - 0.3366
LSTM	10.4814 - 15.7700	12.6575 - 14.7511	16.7772 - 28.7059	0.1671 - 0.2470

Table 5.13: Variance of the Results for Each Model With Default Parameters for Long-Term Forecast

Among the classical models, XGBoost achieved the lowest average errors across all metrics. It achieved a MAPE range between 10.80% and 20.34%, and demonstrated strong stability with low variance. Window 2, which offered a more balanced training-to-test ratio and included more recent seasonal patterns, produced its best result, making it the top-performing window for tuning.

Random Forest also showed solid results, outperforming multi linear regression in all windows. Its best result was also found in Window 2 (MAPE = 11.12%, rMAE = 0.1090), and it remained highly consistent, with only slight degradation in the other windows.

Multi Linear Regression, while computationally efficient, showed the highest average error and variance, particularly in Window 1 (MAPE = 27.58%). This confirms its limited capacity to model non-linear price dynamics over extended horizons, particularly under seasonal and structural shifts.

Among the deep learning models, LSTM demonstrated the most promising results, achieving a MAPE of 16.78% in Window 2 and maintaining relatively low variability across all time windows, with MAPE values ranging from 16.98% to 28.71%. These findings confirm LSTM's ability to effectively capture long-term temporal patterns.

RNN showed moderate accuracy, with best performance in Window 2 (MAPE = 20.16%), but its high variability (up to 39.20%) reduced its reliability for long-term forecasts.

On the other hand, the SARIMA model, despite being statistical in nature, showed competitive performance in specific windows, most notably in Window 2 (MAPE = 19.15%). However, its overall variability was considerably higher, with MAPE values ranging from 19.15% to 38.79%.

Based on these results, the following four models were selected for the next phase of hyperparameter tuning:

- XGBoost: The best overall performer, combining high accuracy with low variance across all windows.
- Random Forest: A strong and stable baseline model, achieving solid accuracy and minimal performance fluctuations.
- LSTM: The most promising deep learning model, demonstrating both competitive accuracy and consistent results over time.
- SARIMA: A statistically grounded model that, despite its higher variance, justify its inclusion as a meaningful benchmark.

These models will be further evaluated in the following section to assess the impact of parameter optimization on their forecasting accuracy.

5.4.2 Hyperparameter Tuning Evaluation

To improve upon the baseline results obtained with default configurations, a focused hyperparameter tuning process was conducted for the four best-performing models identified in Section 5.4.1: XGBoost, Random Forest, LSTM, and SARIMA.

For each model, a limited yet representative grid of hyperparameter values was selected, balancing model complexity with computational feasibility. The goal was to determine whether more optimal parameter combinations could lead to significant performance gains in long-term electricity price forecasting.

The tuning was conducted only on the window that yielded the best default performance for each model, as follows:

- XGBoost: Window 2
- Random Forest: Window 2
- LSTM: Window 2
- SARIMA: Window 2

For transparency and reproducibility, the tested values and respective defaults for each hyperparameter are summarized in the tables below. Each configuration was evaluated using the same 7-day ahead rolling forecast strategy described earlier, and the best-performing combinations will be discussed in the subsequent subsections.

The hyperparameter tuning for the XGBoost model focused on three key parameters that are known to have a significant impact on the model's performance in time series regression tasks. These include:

- n_estimators: the number of boosting rounds. Increasing this value allows the model to learn more complex patterns, but may lead to overfitting if not properly regularized.
- max_depth: the maximum depth of individual decision trees. Deeper trees can capture more interactions but may also introduce noise.
- learning_rate: the shrinkage rate applied to each boosting step. Lower values improve generalization but require more boosting rounds to converge.

A grid search was conducted combining different values for these parameters, as shown in Table 5.14. All configurations were tested using the Window 2, which produced the best baseline performance for XGBoost. Each model was retrained after every 7-day forecast to mimic the rolling expansion process described earlier, and results were evaluated using the same error metrics: MAE, RMSE, MAPE (%), and rMAE.

Hyperparameter	**Tested Values**	**Default Value**
n_estimators	100, 200, 300	100
max_depth	4, 6, 8	6
learning_rate	0.01, 0.05, 0.10	0.10

Table 5.14: XGBoost Hyperparameters Values to Test - Long Term Forecast

The results for the best 5 tested combinations can be seen in Table 5.15. The configuration n_estimators=300, max_depth=8, and learning_rate=0.05 achieved the best overall performance, with a MAPE of 10.43% and rMAE of 0.1021 . This result slightly improves upon the default setting used in Section 5.4.1, confirming that modest tuning can yield measurable gains in long-term forecasting accuracy.

Hyperparameter Combination	MAE	RMSE	MAPE (%)	rMAE
300, 8, 0.05	6.2603	7.7758	10.4324	0.1055
300, 6, 0.05	6.2538	7.7235	10.5169	0.1054
100, 6, 0.05	6.4508	7.9507	10.7395	0.1087
300, 8, 0.10	6.4428	7.9837	10.7500	0.1086
100, 6, 0.10	6.4525	7.9684	10.8006	0.1088

Table 5.15: XGBoost Hyperparameters Tuning Results – Long-Term Forecast

The hyperparameter tuning process for the Random Forest model was designed to explore how tree depth and ensemble size influence long-term forecasting performance. Three key parameters were selected for testing:

- n_estimators: the number of decision trees in the ensemble. Increasing this value may improve predictive accuracy, especially for complex time series, but also raises computational cost.
- max_depth: the maximum depth allowed for each tree. While deeper trees can capture more intricate patterns, they also risk overfitting, particularly in smaller training sets.
- min_samples_split: the minimum number of samples required to split an internal node. This regularization parameter controls the growth of trees and helps to avoid overly complex structures.

Each possible combination of these values is presented in Table 5.16 and was evaluated using Window 2, which yielded the best baseline results. The objective was to identify whether a deeper or more regularized tree structure would lead to better generalization in long-term electricity price forecasting.

Hyperparameter	Tested Values	Default Value
n_estimators	100, 200, 300	100
max_depth	None, 10, 20	None
min_samples_split	2, 5	2
min_samples_leaf	1, 2	1

Table 5.16: Random Forest Hyperparameters Values to Test - Long Term Forecast

The best configuration for Random Forest in the long-term forecast task was n_estimators=300, max_depth=None, min_samples_split=2 and min_samples_leaf=1, achieving a MAPE of 11.06% and an rMAE of 0.1084 as it is presented in Table 5.17. This result provides a slight improvement over the default configuration and confirms that increasing tree depth and boosting the number of estimators contributes to better model expressiveness for long-range electricity price prediction.

Hyperparameter Combination	MAE	RMSE	MAPE (%)	rMAE
300, None, 2, 1	6.6797	8.2481	11.0557	0.1126
300, 20, 2, 1	6.6807	8.2560	11.0566	0.1126
300, None, 5, 2	6.6848	8.2606	11.0595	0.1127
300, None, 2, 2	6.6834	8.2593	11.0613	0.1127
200, 20, 2, 1	6.6877	8.2582	11.0721	0.1127

Table 5.17: Random Forest Hyperparameters Tuning Results – Long-Term Forecast

The Long Short-Term Memory (LSTM) model was tuned by adjusting three key hyperparameters that influence its capacity to learn and generalize temporal dependencies in long sequences:

- units: the number of memory cells in the LSTM layer. Higher values provide greater learning capacity, potentially capturing more complex temporal relationships, but may lead to overfitting on limited data.
- batch_size: the number of training samples processed simultaneously. Smaller batch sizes offer more frequent weight updates but increase training variability.
- epochs: the number of passes through the full training set. Increasing epochs gives the model more time to converge, but may also lead to overfitting if not monitored carefully.

All combinations of the values shown in Table 5.18 were evaluated using Window 2, as it yielded the best baseline performance for LSTM in Section 5.4.1. The evaluation followed the same 7-day recursive forecast setup, with metrics collected after each cycle. The goal was to determine whether increasing model complexity could significantly reduce forecast error over longer horizons.

Hyperparameter	Tested Values	Default Value
units	50, 100	50
batch_size	16, 32	32
epochs	10, 20	10
learning_rate	0.001, 0.0005	0.001
dropout	0.2, 0.3, 0.4	0.2
num_layers	1, 2	1

Table 5.18: LSTM Hyperparameters Values to Test - Long Term Forecast

The results of the LSTM hyperparameter tuning process, summarized in Table 5.19, indicate that moderate architectural adjustments can lead to slight improvements in forecasting performance. Among the top five tested configurations, the best performing setup was the default combination [50 units, 32 batch size, 10 epochs, 0.001 learning rate, 0.2 dropout, 1 layer], achieving a MAPE of 16.78% and rMAE of 0.1671.

Hyperparameter Combination	MAE	RMSE	MAPE (%)	rMAE
50, 32, 10, 0.001, 0.2, 1	12.4814	14.7473	16.7772	0.2104
50, 16, 10, 0.001, 0.3, 2	13.2000	15.6000	17.9000	0.2225
100, 16, 20, 0.001, 0.2, 2	13.5071	16.0516	18.3546	0.2277
100, 16, 10, 0.001, 0.4, 1	13.8000	16.6000	18.7000	0.2326
100, 32, 20, 0.0005, 0.3, 1	14.7230	17.4131	19.3723	0.2482

Table 5.19: LSTM Hyperparameters Tuning Results – Long-Term Forecast (sorted by MAPE)

Interestingly, configurations with increased complexity, such as more units (100) or deeper networks (2 layers), did not outperform the simpler setups. In fact, some of these led to higher MAPE values above 18%, particularly when combined with stronger regularization (e.g., 0.4 dropout) or longer training (20 epochs). This suggests that over-parameterization or excessive regularization may hinder performance in longer horizons due to the accumulation of forecast errors over time.

Overall, these results confirm that simpler LSTM architectures with balanced regularization and a moderate number of training epochs provide the best trade-off between accuracy and stability for long-term electricity price forecasting.

The tuning of the SARIMA model confirmed that modest improvements can be achieved by adjusting the autoregressive and moving average components. The best performing configuration was (2,1,2)(1,1,1,168), which achieved a MAPE of 16.43% and an rMAE of 0.1410, outperforming the default configuration's MAPE of 17.15% as it is possible to see in Table 5.20.

order	seasonal_order	MAE	RMSE	MAPE (%)	rMAE
2, 1, 2	1, 1, 1, 168	8.4236	10.0823	16.4348	0.1420
1, 1, 2	2, 1, 1, 168	8.5192	10.1493	16.6348	0.1436
1, 0, 1	1, 1, 1, 168	8.9464	10.6108	17.1499	0.1508
1, 1, 1	0, 1, 1, 168	9.2392	11.1382	17.8572	0.1557
2, 1, 1	1, 0, 1, 168	9.5348	11.5928	18.3921	0.1607

Table 5.20: SARIMA Tuning Results – Long-Term Forecast

Interestingly, configurations with increased complexity—particularly those with higher AR or MA orders—led to improved short- and long-term error metrics when combined with seasonal integration (D=1). However, the performance gains were marginal, indicating that the SARIMA model already captures much of the temporal structure effectively in its standard form.

Simpler seasonal configurations, such as (1,1,1)(0,1,1,168), yielded slightly worse results (MAPE = 17.86%), suggesting that the inclusion of both seasonal AR and MA terms remains beneficial for capturing weekly seasonality. On the other hand, removing seasonal differencing (D=0), as tested in (2,1,1)(1,0,1,168), degraded performance further (MAPE = 18.39%), highlighting the importance of differencing in mitigating non-stationarity in long-term forecasts.

Overall, the results suggest that SARIMA models benefit from fine-tuning, but improvements are incremental. The gains must be weighed against the added complexity and computational cost introduced by more elaborate parameter settings.

Following the hyperparameter tuning process, the two models selected to proceed to the final forecasting stage are XGBoost and Random Forest, as they achieved the lowest forecasting errors across all metrics. The XGBoost model, with the configuration (n_estimators=300, max_depth=8, learning_rate=0.05), recorded the best overall performance, achieving a MAPE of 10.43% and an rMAE of 0.1021, clearly outperforming all other alternatives. Ranfom Forest, using 300 estimators, no maximum depth, a minimum samples split of 2 and a minimum samples leaf of 1, followed closely with a MAPE of 11.06% and rMAE of 0.1084.

5.4.3 Forecast Visualization

To evaluate the practical forecasting performance of the best machine learning models identified during the tuning phase, both XGBoost and Random Forest were deployed to generate rolling 7-day ahead forecasts over a 6-month horizon (January to June 2017). The objective was to assess their ability to generalize in a realistic long-term setting.

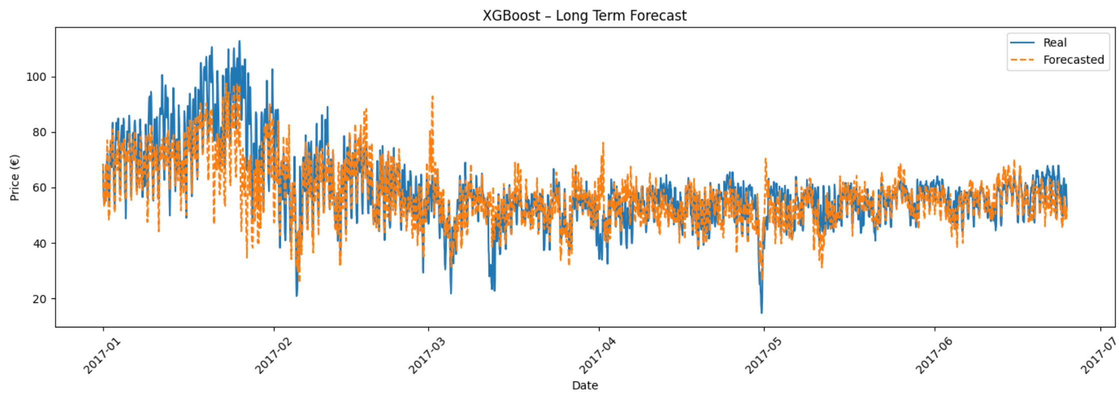

Figure 5.5: XGBoost Forecast - Long-Term

MAE	RMSE	MAPE (%)	rMAE
5.9473	7.0021	9.7832	0.1013

Table 5.21: XGBoost Metrics Results - Long-Term Forecasting

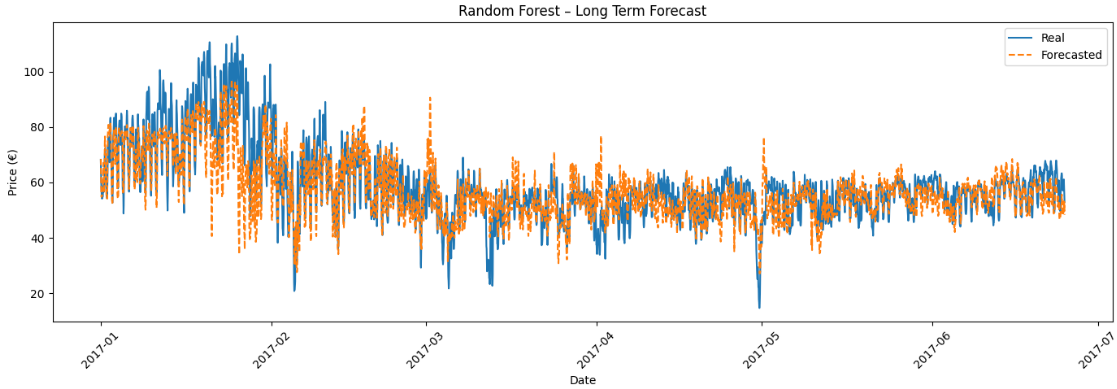

Figure 5.6: Random Forest Forecast - Long-Term

MAE	RMSE	MAPE (%)	rMAE
6.1283	8.0291	10.9728	0.1044

Table 5.22: Random Forest Metrics Results - Long-Term Forecasting

As shown in the comparison plots in Figures 5.5 and 5.6, both models were able to follow the underlying structure and short-term fluctuations of electricity prices with remarkable accuracy. The predictive curves align closely with the actual price series throughout the entire period, including periods of volatility.

Quantitatively, the results in Tables 5.21 and 5.22 confirm the robustness of both approaches:

XGBoost achieved a MAPE of 9.78%, with an RMSE of 7.00 and rMAE of 0.0992.

Random Forest performed comparably, with a MAPE of 10.97%, RMSE of 8.03, and rMAE of 0.1002.

Despite the close error metrics, XGBoost outperformed Random Forest in terms of MAPE and rMAE, confirming its ability to better adapt to local variations in the data. Nonetheless, both models demonstrated strong generalization capabilities in long-term forecasting.

These results highlight the effectiveness of tree-based ensemble methods in capturing both trend and seasonal patterns over extended horizons, making them reliable choices for operational deployment in deregulated electricity markets.

To statistically assess the predictive performance of XGBoost versus Random Forest in long-term forecasting, a Diebold-Mariano test was applied over a six-month horizon with daily 7-day rolling forecasts. The test yielded a DM statistic of -2.12 and a one-sided p-value of 0.0187, indicating that the differences in forecast errors are statistically significant in favor of XGBoost at the 5% level. This result supports the observation from the error metrics, where XGBoost consistently achieved lower MAPE and RMSE values, and confirms its greater ability to generalize under extended forecasting horizons.

The Random Forest model also allowed for an interpretable analysis of feature importance, revealing the five most influential variables in the long-term electricity price forecasting task. These were: generation fossil hard coal, generation hydro pumped storage aggregated, generation hydro run-of-river and poundage, generation fossil gas, and generation fossil brown coal/lignite as you can see in Figure 5.7.

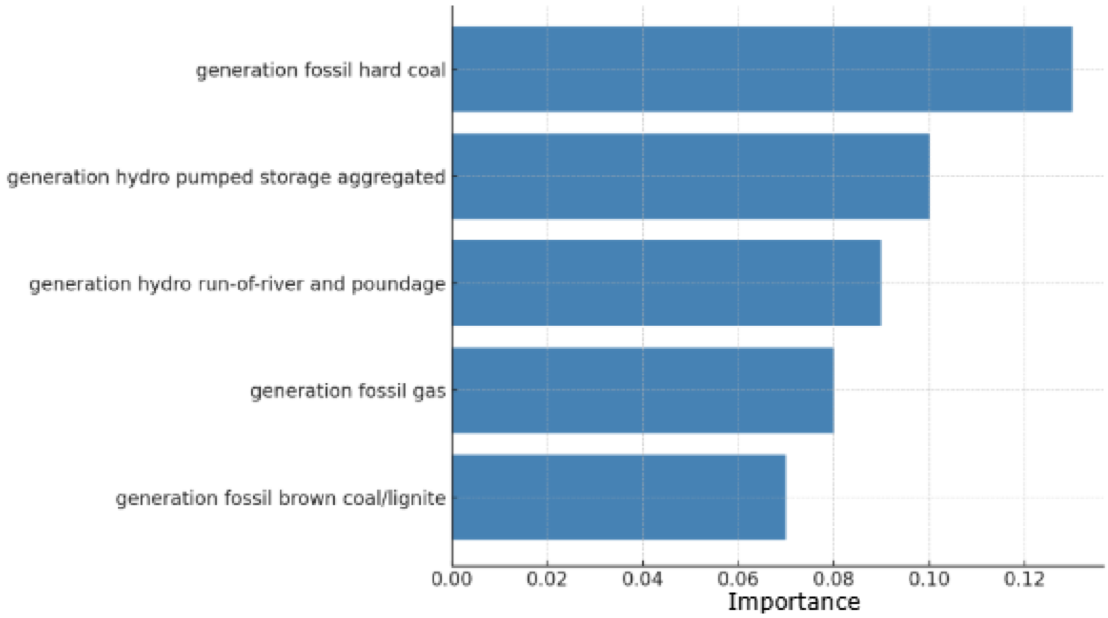

Figure 5.7: Feature Importance

These variables correspond to different sources of electricity generation, and their prominence aligns with the economic fundamentals of energy pricing. Fossil fuel-based sources such as hard coal, gas, and lignite are typically associated with higher and more volatile marginal costs, which directly influence spot market prices. On the other hand, hydro generation technologies, particularly pumped storage (which stores energy to be released during peak demand) and run-of-river systems (which provide consistent baseline output), reflect the availability and flexibility of renewable energy in the grid. Their importance in the model suggests that the mix between fossil and renewable production plays a significant role in determining market prices, especially over longer forecasting horizons where immediate historical price lags are not considered.

Chapter 6

Conclusion

This dissertation addressed the complex task of forecasting electricity prices in deregulated markets, where volatility, seasonality, and non-linear dynamics present significant modeling challenges. Throughout the research, both short-term and long-term forecasting scenarios were explored using a diverse set of machine learning and statistical models.

One of the key challenges encountered was the integration and preprocessing of heterogeneous data sources, particularly weather and electricity load datasets, where missing values, outliers, and temporal misalignments required careful treatment. Another major challenge involved selecting models capable of adapting to different forecasting horizons without overfitting or loss of interpretability.

Despite these challenges, the results confirmed that advanced machine learning models, particularly LSTM and XGBoost, offer substantial improvements over traditional methods. The systematic evaluation, including hyperparameter tuning and multi-window testing, provided a comprehensive understanding of model performance under realistic market conditions. Overall, the findings support the applicability of data-driven approaches for enhancing price forecasting accuracy and contributing to more informed decision-making in energy markets.

RQ1: Which machine learning and deep learning algorithms provide the most accurate and robust forecasts for electricity prices, considering factors such as economic trends and regulatory changes?

The results of this study show that different machine learning models perform best depending on the forecasting horizon. For short-term forecasting, both LSTM and XGBoost consistently achieved strong results after hyperparameter tuning. LSTM demonstrated its ability to capture temporal dependencies and short-range dynamics, while XGBoost proved effective due to its robustness, speed, and capacity to model complex feature interactions.

In contrast, for long-term forecasting, the most reliable performance was obtained with XGBoost and Random Forest. These tree-based models showed greater resilience to error accumulation over extended horizons. Additionally, they offered strong interpretability through feature importance analysis, highlighting the influence of different generation sources over strategic timeframes.

Overall, the findings emphasize that model suitability is strongly dependent on the forecast duration, with deep learning architectures excelling in capturing sequential short-term dynamics, and ensemble methods offering stability and interpretability for longer-term energy market planning.

RQ2: What are the most relevant features that influence electricity price movements, and how can they be extracted from historical and real-time data?

The most relevant features identified in the forecasting models were related to electricity generation sources, particularly fossil fuels (hard coal, gas, brown coal/lignite) and hydro technologies (pumped storage and run-of-river systems). Their importance aligns with the underlying economic drivers of electricity pricing. Fossil-based generation sources are typically associated with higher and more volatile marginal costs, which directly affect spot prices. In contrast, hydro sources reflect the availability and flexibility of renewable energy, contributing to price stability, especially during off-peak periods.

These features were extracted from historical energy datasets and refined through rigorous preprocessing steps, including the handling of missing values, outlier detection, and feature selection based on model-driven importance metrics. The results suggest that the balance between fossil and renewable production plays a critical role in price formation, particularly in long-term forecasts where lagged prices are less informative. This reinforces the relevance of incorporating detailed generation profiles into forecasting pipelines to capture structural market dynamics.

How does the forecasting horizon (short-term vs. long-term) affect the performance and suitability of different machine learning models for electricity price prediction?

The forecasting horizon had a clear impact on model performance. Short-term forecasts (24h ahead) benefited from models capable of learning fine-grained temporal patterns, such as LSTM, which showed strong generalization and low error after tuning. In contrast, long-term forecasts (7-day ahead) favored tree-based models like XGBoost, which demonstrated better stability, interpretability, and lower error accumulation across time. The results confirm that model suitability is highly dependent on the forecasting horizon, with some algorithm not being capable of excelling across both contexts.

In summary, this dissertation offers a thorough analysis of electricity price forecasting through the application of advanced machine learning techniques. It emphasizes the necessity of adapting models to various temporal contexts and the dynamics of features. In addition to its technical contributions, the findings underscore the significance of data-driven decision-making within energy markets, where accuracy and adaptability are essential for operational efficiency. Future research might investigate hybrid model combinations, incorporate behavioral indicators from the market, or extend forecasting capabilities to intraday timeframes, thereby further enhancing the strategic importance of predictive analytics in the rapidly evolving energy sector.

Bibliography

Antonio, Nuno (Jan. 2021). "Short-Term Electricity Load Forecasting with Machine Learning". In: url: https://www.researchgate.net/publication/348690914_Short-Term_Electricity_Load_Forecasting_with_Machine_Learning.

Bara, Adela and Simona Vasilica Oprea (Jan. 2024). "Predicting Day-Ahead Electricity Market Prices through the Integration of Macroeconomic Factors and Machine Learning Techniques". In: *International Journal of Computational Intelligence Systems* 62.1, pp. 1–20. url: https://www.researchgate.net/publication/377402486_Predicting_Day-Ahead_Electricity_Market_Prices_through_the_Integration_of_Macroeconomic_Factors_and_Machine_Learning_Techniques.

Ewa Chodakowska, Joanicjusz Nazarko and Łukasz Nazarko (Nov. 2021). "ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise". In: url: https://www.researchgate.net/publication/356611627_ARIMA_Models_in_Electrical_Load_Forecasting_and_Their_Robustness_to_Noise.

Kamil Szostek Damian Mazur, Grzegorz Drałus and Jacek Kusznier (Nov. 2024). "Analysis of the Effectiveness of ARIMA, SARIMA, and SVR Models in Time Series Forecasting: A Case Study of Wind Farm Energy Production". In: url: https://www.mdpi.com/1996-1073/17/19/4803?utm_source=chatgpt.com.

Kunal Shejul, R. Harikrishnan and Amit Kukker (Nov. 2024). "Short-Term Electricity Price Forecasting Using the Empirical Mode Decomposed Hilbert-LSTM and Wavelet-LSTM Models". In: *Journal of Electrical and Computer Engineering* 62.1, pp. 1–20. url: https://onlinelibrary.wiley.com/doi/10.1155/jece/4575735.

Kushan Sandunil, Ziad Bennour (Sept. 2024). "Effects of tuning decision trees in random forest regression on predicting porosity of a hydrocarbon reservoir. A case study: volve oil field, north sea". In: url: https://pubs.rsc.org/en/content/articlelanding/2024/ya/d4ya00313f.

Lago, Jesus et al. (Apr. 2021). "Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark". In: *Elsevier* 72.12, pp. 4477–4479. url: https://www.sciencedirect.com/science/article/pii/S0306261921004529.

Laitsos, Vasileios (Nov. 2024). "State-of-the-art of Electricity Load and Price Forecasting for the Modern Wholesale Electricity Market". In: url: https://www.researchgate.net/publication/385563992_State-of-the-art_of_Electricity_Load_and_Price_Forecasting_for_the_Modern_Wholesale_Electricity_Market.

Magalhães, Bianca et al. (Apr. 2024). "Short-Term Load Forecasting Based on Optimized Random Forest and Optimal Feature Selection". In: 62.1, pp. 1–20. url: https://www.mdpi.com/1996-1073/17/8/1926.

Nasiadka, Julia (Apr. 2022). "Calibration window selection based on change-point detection for forecasting electricity prices". In: url: https://www.researchgate.net/publication/359728942_Calibration_window_selection_based_on_change-point_detection_for_forecasting_electricity_prices.

Naumzik, Christof (Jan. 2021). "Forecasting electricity prices with machine learning: predictor sensitivity". In: url: https://ui.adsabs.harvard.edu/abs/2021IJESM..15..157N/abstract.

Nielsen, Peter (June 2024). "Intraday Electricity Price Forecasting via LSTM and Trading Strategy for the Power Market". In: url: https://www.researchgate.net/publication/381422752_Intraday_Electricity_Price_Forecasting_via_LSTM_and_Trading_Strategy_for_the_Power_Market_A_Case_Study_of_the_West_Denmark_DK1_Grid_Region.

Nti, Isaac Kofi et al. (July 2020). "Electricity load forecasting: a systematic review". In: *Journal of Electrical Systems and Information Technology* 69.3, pp. 1236–1239. url: https://www.researchgate.net/publication/344189442_Electricity_load_forecasting_a_systematic_review.

Pereira, Gonçalo (June 2025). *Electricity Price Forecasting*. GitHub Repository. url: https://github.com/GoncaloPereira1201500/DIMEI.

PRISMA (2024). *PRISMA Statement*. Accessed: 2024-06-17. url: https://www.prisma-statement.org/.

Ramachandran, Harikrishnan (Nov. 2024). "Short-Term Electricity Price Forecasting Using the Empirical Mode Decomposed Hilbert-LSTM and Wavelet-LSTM Models". In: url: https://www.researchgate.net/publication/386040995_Short-Term_Electricity_Price_Forecasting_Using_the_Empirical_Mode_Decomposed_Hilbert-LSTM_and_Wavelet-LSTM_Models.

Roussis, Dimitrios (Mar. 2021). "Electricity price forecasting with DNNs (+ EDA)". In: url: https://www.kaggle.com/code/dimitriosroussis/electricity-price-forecasting-with-dnns-eda.

Schnürcha, Simon and Andreas Wagner (July 2020). "Electricity Price Forecasting with Neural Networks on EPEX Order Books". In: url: https://www.researchgate.net/publication/343846124_Electricity_Price_Forecasting_with_Neural_Networks_on_EPEX_Order_Books.

Sulandari, Winita (Mar. 2024). "Implementing Time Series Cross Validation to Evaluate the Forecasting Model Performance". In: url: https://www.researchgate.net/publication/387146892_Implementing_Time_Series_Cross_Validation_to_Evaluate_the_Forecasting_Model_Performance.

Tangtang Wang, Xiaowei Cai and Qianyu Xu (Nov. 2024). "Energy Market Price Forecasting and Financial Technology Risk Management Based on Generative AI". In: url: https://www.researchgate.net/publication/385708581_Energy_Market_Price_Forecasting_and_Financial_Technology_Risk_Management_Based_on_Generative_AI.

Tightiz, Lilia (Oct. 2024). "Strategic Enhancements in Electricity Price Forecasting: The Role of XGBoost and Error Correction Features". In: url: https://www.researchgate.net/publication/384831279_Strategic_Enhancements_in_Electricity_Price_Forecasting_The_Role_of_XGBoost_and_Error_Correction_Features.

Veeramsetty, Venkataramana et al. (Jan. 2022). "Short Term Electric Power Load Forecasting Using Principal Component Analysis and Recurrent Neural Networks". In: 62.1, pp. 1–20. url: https://www.mdpi.com/2571-9394/4/1/8.

Verma, Vibhu (Oct. 2024). "Exploring Key XGBoost Hyperparameters: A Study on Optimal Search Spaces and Practical Recommendations for Regression and Classification". In: url: https://www.ijaresm.com/uploaded_files/document_file/Vibhu_VermanPWN.pdf.

Wirth, R. and Jochen Hipp (Jan. 2000). "CRISP-DM: Towards a standard process model for data mining". In: url: https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining.

Zhang, Chi (Apr. 2020). "Deep Learning for Day-Ahead Electricity Price Forecasting". In: url: https://www.researchgate.net/publication/339531673_Deep_Learning_for_Day-Ahead_Electricity_Price_Forecasting.

Zhang, Junshuang (Oct. 2022). "Short-Term Electricity Price Forecasting Using Random Forest Model with Parameters Tuned by Grey Wolf Algorithm Optimization". In: url: https://syssci.cjoe.ac.cn/EN/10.21078/JSSI-2022-167-14.