



MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MÁSTER

VARIABLES QUE IMPACTAN EN EL MERCADO DE EMISIONES
DE CARBONO DE LA UE: PREDICCIONES CON MODELOS
ESTADÍSTICOS Y DE MACHINE LEARNING

Autor: Ignacio de Vicente-Tutor Vázquez
Director: José Carlos Romero Mora

Madrid
Enero 2025

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
Variables que impactan en el mercado de emisiones de carbono de la UE: predicciones con modelos
estadísticos y de machine learning

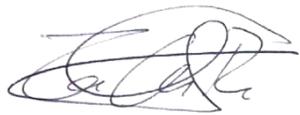
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el
curso académico 2024/25 es de mi autoría, original e inédito y
no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido
tomada de otros documentos está debidamente referenciada.

Fdo.:  Fecha: 14/ 01/ 2025

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.:  Fecha: 14/ 01/ 2025



MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MÁSTER

VARIABLES QUE IMPACTAN EN EL MERCADO DE EMISIONES
DE CARBONO DE LA UE: PREDICCIONES CON MODELOS
ESTADÍSTICOS Y DE MACHINE LEARNING

Autor: Ignacio de Vicente-Tutor Vázquez
Director: José Carlos Romero Mora

Madrid
Enero 2025

Variables que impactan en el mercado de emisiones de carbono de la UE: predicciones con modelos estadísticos y de machine learning

Autor: Ignacio de Vicente-Tutor Vázquez

Director: José Carlos Romero Mora

Entidad colaboradora: ICAI – Universidad Pontificia de Comillas

Resumen del proyecto

Este proyecto tiene como objetivo analizar las relaciones entre variables clave del mercado de carbono de la Unión Europea y desarrollar modelos predictivos basados en estas relaciones. La investigación se centra en identificar y seleccionar las variables más relevantes mediante métodos de análisis de correlación, reducción de dimensionalidad y evaluación de multicolinealidad. A partir de este análisis, se implementan diversos modelos predictivos, incluyendo algoritmos de regresión y redes neuronales avanzadas como el Perceptrón Multicapa (MLP) y las Redes LSTM (Long Short-Term Memory).

El trabajo explora el desempeño de estos modelos para predecir la evolución de los precios de las emisiones de CO₂ en el marco del Sistema de Comercio de Emisiones de la Unión Europea (EU ETS). Asimismo, se compara la efectividad de las técnicas de modelado tradicionales frente a enfoques basados en aprendizaje profundo, proporcionando conclusiones sobre la utilidad práctica de cada metodología.

La relevancia del proyecto radica en su contribución al entendimiento del mercado de carbono y en el desarrollo de herramientas que pueden apoyar la toma de decisiones en sectores clave de la economía sostenible.

Palabras Clave: carbono, predicción, aprendizaje automático, correlación, modelos predictivos

1. Introducción

El mercado de emisiones de carbono de la Unión Europea (UE) es un sistema clave en la lucha contra el cambio climático, diseñado para reducir de manera eficiente y efectiva las emisiones de gases de efecto invernadero. Este sistema, conocido como el Sistema de Comercio de Emisiones de la UE (EU ETS, por sus siglas en inglés), opera bajo el principio de "cap and trade", estableciendo un límite (cap) a las emisiones permitidas y permitiendo que las empresas compren y vendan permisos (trade) para emitir dióxido de carbono (CO₂). Este enfoque proporciona un incentivo económico para reducir las emisiones, fomentando la innovación y la adopción de tecnologías más limpias.

Sin embargo, el mercado de emisiones de carbono es un sistema complejo que depende de múltiples variables económicas, energéticas y políticas. Factores como los precios de la energía, las políticas gubernamentales, la actividad económica y la transición hacia fuentes de energía renovables influyen directamente en su funcionamiento y efectividad. Dado este contexto, resulta crucial analizar cómo estas variables interactúan y afectan el precio de las emisiones de carbono, ya que esto tiene implicaciones directas en la toma de decisiones de empresas y gobiernos, así como en la transición hacia una economía más sostenible.

Este trabajo busca explorar estas interacciones utilizando modelos estadísticos y técnicas avanzadas de machine learning. Estas herramientas permiten analizar grandes volúmenes de datos y detectar patrones complejos que serían difíciles de identificar mediante métodos tradicionales. A través de este enfoque, se pretende no solo comprender mejor las dinámicas del mercado de emisiones de carbono, sino también ofrecer predicciones más precisas que puedan servir como base para la formulación de políticas y estrategias empresariales. Como veremos a continuación, este análisis abarca una selección cuidadosa de variables relevantes y la implementación de modelos que capturen tanto relaciones lineales como no lineales en el mercado de emisiones de carbono.

2. Objetivos del trabajo

El proyecto de fin de máster tiene como finalidad abordar varios aspectos fundamentales del Sistema de Comercio de Emisiones de la Unión Europea (EU ETS), mediante el análisis detallado

y la predicción de la dinámica del mercado de derechos de emisión. A continuación, se presentan los cinco objetivos principales que persigue resolver:

1. Comprender la dinámica del mercado de derechos de emisión:

El primer objetivo consiste en realizar un análisis exhaustivo de las variables clave que influyen en el mercado de derechos de emisión, como los precios de los derechos, las emisiones por sector y país, y los indicadores macroeconómicos. Este análisis proporcionará una comprensión profunda de las dinámicas fundamentales que rigen el mercado del EU ETS, permitiendo identificar las principales fuerzas impulsoras y desafíos.

2. Analizar la correlación entre variables económicas y políticas con el mercado de derechos de emisión:

Se busca explorar cómo las variables económicas y políticas, incluidas las decisiones regulatorias y las condiciones macroeconómicas, afectan los precios y las cantidades de los derechos de emisión. Identificar estas correlaciones es esencial para entender la sensibilidad del mercado a diferentes factores y cómo estos pueden ser manipulados o anticipados para lograr los objetivos de reducción de emisiones.

3. Desarrollar modelos predictivos para cambios significativos en el mercado:

El tercer objetivo se centra en la construcción de modelos predictivos que estimen la probabilidad de cambios importantes en el mercado, utilizando variables históricas. Esto incluye la aplicación de técnicas estadísticas y de machine learning para predecir la dirección y magnitud de los movimientos del mercado, lo que es crucial para la toma de decisiones estratégicas en la gestión de emisiones.

3. Definición del Proyecto

El trabajo se estructura en varias partes fundamentales, cada una diseñada para abordar aspectos clave del análisis y modelado del mercado de emisiones de carbono en la Unión Europea. La primera parte presenta una introducción al contexto general del mercado de emisiones, describiendo su relevancia y los principales factores que lo afectan. La segunda parte se centra en la base teórica que sustenta el análisis, incluyendo una revisión de la literatura y de los modelos estadísticos y de machine learning empleados. La tercera parte detalla el proceso de recopilación,

selección y preparación de las variables utilizadas en el estudio. La cuarta parte se dedica al desarrollo de los modelos predictivos, describiendo las técnicas aplicadas, los procesos de entrenamiento y validación, y los resultados obtenidos. La quinta parte ofrece una comparación entre los distintos modelos implementados, analizando sus capacidades predictivas y la precisión de sus resultados. Por último, la sexta parte incluye la bibliografía utilizada, que respalda y contextualiza los análisis realizados.

Se otorga especial relevancia a la cuarta parte, ya que constituye el núcleo del proyecto. Esta sección comienza con una descripción detallada de las herramientas y metodologías utilizadas para construir los modelos predictivos, destacando las diferencias entre los enfoques estadísticos y de machine learning. Posteriormente, se aborda el proceso de selección de variables mediante análisis de correlaciones y técnicas de reducción de dimensionalidad, asegurando que los modelos se centren en las variables más relevantes. A continuación, se presenta el desarrollo e implementación de los modelos, analizando cómo se entrenaron y evaluaron para garantizar su robustez y precisión. Finalmente, se exponen los resultados obtenidos y su interpretación en el contexto del mercado de emisiones de carbono.

El análisis se basa en un conjunto de datos que incluye información económica, energética y financiera relevante, combinada con los precios de las emisiones de carbono en la Unión Europea. Las técnicas de modelado empleadas abarcan desde métodos tradicionales, como regresiones lineales, hasta enfoques más avanzados, como redes neuronales y modelos de memoria a largo plazo. Este enfoque integral permite capturar tanto relaciones lineales como no lineales en los datos, ofreciendo una visión completa y precisa de las dinámicas del mercado de emisiones de carbono.

4. Resultados

El estudio realizado consistió en el análisis de diferentes modelos predictivos aplicados al precio del carbono en el mercado EU ETS. Estos modelos incluyeron enfoques lineales, métodos basados en árboles de decisión y redes neuronales avanzadas. El objetivo principal fue evaluar la precisión y capacidad de cada modelo para capturar la complejidad y volatilidad de este mercado, caracterizado por su dependencia de múltiples factores económicos, energéticos y regulatorios.

Los modelos lineales, como la regresión múltiple y polinómica, mostraron limitaciones significativas, reflejadas en sus altos errores de predicción (RMSE) y coeficientes de determinación negativos. Aunque las técnicas de regularización como Lasso y Ridge mejoraron levemente el desempeño, estos enfoques resultaron insuficientes para modelar la naturaleza compleja de los datos. Por su parte, los métodos basados en árboles, incluidos Random Forest, Gradient Boosting y XGBoost, ofrecieron un mejor ajuste, capturando relaciones no lineales con mayor precisión. Sin embargo, estos modelos tampoco lograron explicar de manera adecuada la variabilidad observada, manteniéndose con R^2 negativos en todas sus configuraciones.

En contraste, las redes neuronales, especialmente las Long Short-Term Memory (LSTM), destacaron en el análisis. La capacidad de las LSTM para manejar relaciones temporales permitió obtener predicciones notablemente más precisas. En su configuración optimizada, este modelo alcanzó un RMSE de 0.4408 y un R^2 de 0.9757, lo que evidencia su capacidad superior para capturar la dinámica del mercado del carbono. Aunque el modelo Multilayer Perceptron (MLP) optimizado mostró mejoras importantes frente a métodos lineales y de ensamble, no logró igualar el desempeño de las LSTM.

Los resultados indican que los enfoques tradicionales tienen limitaciones para modelar un problema tan complejo como el precio del carbono, mientras que las redes neuronales avanzadas ofrecen una solución más precisa y robusta para esta tarea.

5. Conclusiones

El análisis comparativo de los modelos predictivos ha permitido identificar fortalezas y limitaciones en el contexto del mercado EU ETS, donde la volatilidad y la interacción de múltiples factores hacen imprescindible el uso de enfoques sofisticados. Si bien los modelos lineales y basados en árboles ofrecieron una perspectiva inicial sobre las dinámicas del mercado, su desempeño fue insuficiente para capturar de manera efectiva las complejidades temporales y no lineales que caracterizan este problema.

Por otro lado, el empleo de redes neuronales avanzadas, particularmente las LSTM, resaltó como una estrategia significativamente superior. Estos modelos no solo permitieron reducir de manera

notable los errores de predicción, sino que además demostraron una capacidad única para integrar dependencias temporales y capturar fluctuaciones complejas en los datos. Este avance no solo posiciona a las LSTM como la herramienta más adecuada para este tipo de tareas, sino que también subraya la importancia de incorporar metodologías adaptadas a la naturaleza específica de los datos analizados.

En conjunto, este estudio reafirma que, frente a la complejidad inherente del mercado EU ETS, las soluciones tradicionales son limitadas y que las innovaciones en aprendizaje automático y redes neuronales pueden proporcionar un marco predictivo más sólido y eficaz. Estos hallazgos no solo tienen implicaciones para la predicción del precio del carbono, sino que también abren nuevas oportunidades para abordar desafíos similares en otros mercados complejos y dinámicos.

6. Referencias

- [1] D. Ellerman, F. J. Convery, and C. de Perthuis. Pricing Carbon: The European Union Emissions Trading Scheme. Cambridge University Press, 2010.
- [2] Organisation for Economic Co-operation and Development (OECD). Carbon Pricing: What Role for Border Carbon Adjustments. Technical Report, 2021.
- [3] Journal of Cleaner Production. The Impact of Carbon Price and Firm-Level Emissions on Swedish Firms' Profits. 2023.
- [4] Energy Economics. The Environmental-Financial Performance Nexus of EU ETS Firms: A Quantile Regression Approach. 2022.
- [5] K. S. Rogge, M. Schneider, and V. H. Hoffmann. The Innovation Impact of the EU Emission Trading System: Findings of Company Case Studies in the German Power Sector. Ecological Economics, 70(2011), 513–523.
- [6] BI Norwegian Business School. EU ETS & the Maritime Shipping Industry: Barriers Compromising the Economic Efficiency and Effectiveness of the System. 2024.
- [7] Cambridge Press. The European Union's Emissions Trading System in Perspective. 2019.
- [8] R. Rafaty, G. Dolphin, and F. Pretis. Carbon Pricing and the Elasticity of CO₂ Emissions. Energy Policy Research Group, University of Cambridge, Working Paper 2035, 2020.
- [9] A. Wach. The European Union Emissions Trading Scheme: Failure Analysis and Assessment of Market Stability Reserve Solution. Norwegian School of Economics, Bergen, 2016.
- [10] European Commission. The EU Emissions Trading System. European Commission, 2021.

- [11] L. H. Goulder and A. R. Schein. Carbon Pricing in the EU: Evaluation of Different EU ETS Reform Options. Technical Report, 2022.
- [12] C. Flachsland, N. Koch, and S. Fuss. A Framework for Assessing the Performance of Cap-and-Trade Systems: Insights from the EU ETS. *Review of Environmental Economics and Policy*, 2018.
- [13] M. A. Mehling and R. A. Ritz. Emissions Trading Hybrids: The Case of the EU ETS. *Energy Economics*, 2022.
- [14] European Commission, Climate Action, EU ETS Emissions Cap. [En línea]. Disponible en: https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/eu-ets-emissions-cap_en.
- [15] M. Cames and M. Frerk. The Best (and Worst) of GHG Emission Trading Systems: Comparing the EU ETS with Its Followers. *Journal of Environmental Policy*, 2019.
- [16] R. Martin, M. Muûls, and U. J. Wagner. Allocation and Competitiveness in the EU Emissions Trading Scheme: Policy Overview. Carbon Markets Series, 2020.
- [17] F. Laurent. The European Ships' Monitoring, Reporting and Verification (MRV): Pre-evaluation of a Regional Regulation on Carbon Dioxide Inventory. IAME Conference, 2016.
- [18] J. Verschuuren and F. Fleurke. Report on the Legal Implementation of the EU ETS at Member State Level. Tilburg University, 2014.
- [19] J. Kruger and C. Egenhofer. Confidence through Compliance in Emissions Trading Markets. *Sustainable Development Law & Policy*, 2005.
- [20] T. Santier. Monitoring, Reporting, and Verification in a Carbon Border Adjustment Mechanism. 2023.
- [21] M. Zhang, Y. Liu, and Y. Su. Comparison of Carbon Emission Trading Schemes in the European Union and China. *Climate*, 2017.
- [22] P. Gailhofer and V. Graichen. Monitoring, Reporting and Verification in the EU ETS: Legal Design Considerations. 2024.
- [23] C. Warnecke. Can CDM Monitoring Requirements Be Reduced While Maintaining Environmental Integrity?. *Climate Policy*, 2014.
- [24] F. Deane, A. Huggins, and M. S. Karim. Measuring, Monitoring, Reporting and Verification of Shipping Emissions. *Review of European, Comparative & International Environmental Law*, 2019.
- [25] K. Kulovesi, E. Mekkonen, y T. Voigt. Enforcement of the EU ETS in the Member States. *Journal of Environmental Law*, 2021.
- [26] Red Nacional de Derechos de Emisión (RENADE), Taxes Renade. [En línea]. Disponible en: <https://www.renade.es/ing/Information/Taxes-Renade>.

- [27] J. Schleich, R. Betz, y K. Rogge. EU Emissions Trading – Better Job Second Time Around. European Environment, 2007.
- [28] J. Schleich, R. Betz, and K. Rogge. EU Emissions Trading – Better Job Second Time Around?. European Environment, 17(5), 279–294, 2007.
- [29] J. Wettestad and T. Jevnaker. Rescuing EU Emissions Trading: The Climate Policy Flagship. Springer, 2016.
- [30] J. B. Skjærseth and J. Wettestad. EU Emissions Trading: Initiation, Decision-Making and Implementation. Global Environmental Politics, 8(1), 117–141, 2008.
- [31] European Commission. The EU Emissions Trading System (EU ETS). Technical Report, European Commission, 2021.
- [32] G. Perino y M. Willner. Procrastinating Reform: The Impact of the Market Stability Reserve on the EU ETS. Journal of Environmental Economics and Management, 2016.
- [33] C. Flachsland, R. Marschinski, and O. Edenhofer. To Link or Not to Link: Benefits and Disadvantages of Linking Cap-and-Trade Systems. Climate Policy, 9(4), 358–372, 2009.
- [34] J. Köhler, M. Grubb, and D. Popp. Transition Pathways and Risk Analysis for Climate Change Mitigation. Climate Policy, 19(3), 342–358, 2019.
- [35] M. Duwe, J. Graichen, y H. Böttcher. Can Current EU Climate Policy Reliably Achieve Climate Neutrality by 2050?. Ecologic Institute, 2023.
- [36] J. Boratyński, M. Pyrka, S. Lizak, I. Tobiasz, y R. Jeszke. Reform of the Market Stability Reserve (MSR) in the “Fit for 55” Package. IOS-PIB, 2022.
- [37] M. Wiels and J. C. Defraigne. The EU's Climate Ambitions: Assessment of the Social Cost of the Fit for 55 Package. Université Catholique de Louvain, 2023.
- [38] J. C. Richstein, É. J. L. Chappin, y L. J. de Vries. The Market (In-)Stability Reserve for EU Carbon Emission Trading: Why It Might Fail and How to Improve It. Utilities Policy, 2015.
- [39] D. Seibert, P. Kasten, J. Graichen, y N. Wissner. EU 2040 Climate Target: Contributions of the Transport Sector. Oeko Institute, 2024.
- [40] H. Rienks y B. Moore. EU Innovation Funding for Climate Neutrality. Wageningen University, 2023.
- [41] T. Hesse y S. Braungardt. EU 2040 Climate Target: Contributions of the Buildings Sector. Oeko Institute, 2024.
- [42] B. Chen and R. Wu. Comparative Analysis of the Carbon Market in the EU and China. Journal of Environmental Policy, 2023.
- [43] F. Annunziata. Emission Allowances as Financial Instruments. In Governance, Financial Stability and Financial Markets, Springer, 2021.

- [44] A. Michaelowa and I. Shishlov. Evolution of International Carbon Markets: Lessons for the Paris Agreement. *Wiley Interdisciplinary Reviews: Climate Change*, 10(6), e613, 2019.
- [45] M. Li and M. Duan. Exploring Linkage Opportunities for China's Emissions Trading System Under the Paris Targets: EU-China and Japan-Korea-China Cases. *Energy Economics*, 2021.
- [46] World Bank, Carbon Pricing Dashboard. [En línea]. Disponible en: <https://carbonpricingdashboard.worldbank.org/>.
- [47] C. Müller. Rethinking Voluntary Carbon Credit Markets and the EU ETS: A Critical Review. Fulda University of Applied Sciences, 2024.
- [48] D. Tänzler, J. Groß, and C. Warnecke. Interactions Between New Market Mechanisms and Emissions Trading Schemes: Opportunities and Prospects for Countries to Use Article 6 of the Paris Agreement. German Federal Ministry for the Environment, 2018.
- [49] O. M. Hladka, I. M. Karpovich, and A. Y. Sabaniuk. Artificial Neural Networks for Analysis of Stock Market Dynamics: Applications of Python and Jupyter Notebooks. *Vernadsky Tech Journals*, 2024.
- [50] E. Champa-Bujaico, A. M. Díez-Pascual, A. Lomas Redondo, and P. Garcia-Diaz. Optimization of mechanical properties of multiscale hybrid polymer nanocomposites: A combination of experimental and machine learning techniques. *Composites Part B*, vol. 269, p. 111099, 2024.
- [51] R. Bhattacharjee, K. Botchway, J. C. Pashin, G. Chakraborty, and P. Bikina, "Machine learning-based prediction of CO₂ fugacity coefficients: Application to estimation of CO₂ solubility in aqueous brines as a function of pressure, temperature, and salinity," *International Journal of Greenhouse Gas Control*, vol. 128, p. 103971, 2023.
- [52] T. Souza, J. Herrera, A. Johnson, and K. B. Wright. Detection and classification of pneumonia using the Orange3 data mining tool. *International Journal of Computational Intelligence Systems*, vol. 14, no. 3, pp. 250-263, 2023.
- [53] F. Cheng, C. Yang, C. Zhou, L. Lan, H. Zhu, and Y. Li. Simultaneous Determination of Metal Ions in Zinc Sulfate Solution Using UV–Vis Spectrometry and SPSE-XGBoost Method. *Sensors*, vol. 20, no. 17, p. 4936, 2020.
- [54] K. Y. Chan, B. Abu-Salih, R. Qaddoura, A. M. Al-Zoubi, V. Palade, D.-S. Pham, J. Del Ser, and K. Muhammad. "Deep neural networks in the cloud: Review, applications, challenges and research directions." *Neurocomputing*, vol. 545, p. 126327, 2023.
- [55] B. Yang, Z. Guo, L. Wang, J. He, B. Xia, and S. Vakily. Updated Global Navigation Satellite System Observations and Attention-Based Convolutional Neural Network–Long Short-Term Memory Network Deep Learning Algorithms to Predict Landslide Spatiotemporal Displacement. *Remote Sens.*, vol. 15, no. 20, p. 4971, 2023.

Variables impacting the EU carbon emission market: predictions with statistical models and machine learning

Author: Ignacio de Vicente-Tutor Vázquez

Advisor: José Carlos Romero Mora

Collaborating entity: ICAI – Universidad Pontificia de Comillas

Project summary

This project aims to analyze the relationships between key variables in the European Union carbon market and develop predictive models based on these relationships. The research focuses on identifying and selecting the most relevant variables through correlation analysis, dimensionality reduction techniques, and multicollinearity evaluation. Using this analysis as a foundation, various predictive models are implemented, including regression algorithms and advanced neural networks such as Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) networks.

The project explores the performance of these models in forecasting CO₂ emission price trends within the European Union Emissions Trading System (EU ETS). Furthermore, it compares the effectiveness of traditional modeling techniques against deep learning approaches, providing insights into the practical utility of each methodology.

The project's relevance lies in its contribution to understanding the carbon market and developing tools to support decision-making in key sectors of a sustainable economy.

Keywords: carbon, prediction, machine learning, correlation, predictive models

1. Introduction

The European Union carbon emissions market is a key system in the fight against climate change, designed to efficiently and effectively reduce greenhouse gas emissions. Known as the EU Emissions Trading System (EU ETS), it operates on a "cap and trade" principle, setting a limit (cap) on allowable emissions and enabling companies to buy and sell allowances (trade) for

emitting carbon dioxide (CO₂). This mechanism provides an economic incentive to reduce emissions, encouraging innovation and the adoption of cleaner technologies.

However, the carbon market is a complex system influenced by various economic, energy-related, and policy factors. Elements such as energy prices, governmental policies, economic activity, and the shift toward renewable energy sources directly impact its functionality and effectiveness. In this context, understanding how these variables interact and influence carbon emission prices is crucial for informed decision-making by businesses and governments, as well as for transitioning to a more sustainable economy.

This study seeks to explore these interactions using statistical models and advanced machine learning techniques. These tools enable the analysis of large datasets and uncover complex patterns that are difficult to identify with traditional methods. By adopting this approach, the research aims not only to gain deeper insights into the dynamics of the carbon emissions market but also to provide more accurate predictions that can inform policy-making and business strategies.

As detailed in the subsequent sections, this analysis involves careful selection of relevant variables and the implementation of models that capture both linear and nonlinear relationships within the carbon emissions market.

2. Objectives of the thesis

The master's thesis aims to address several fundamental aspects of the European Union Emissions Trading System (EU ETS) through a detailed analysis and prediction of the dynamics of the emission allowances market. Below are the three main objectives the project seeks to achieve:

1. Understand the dynamics of the emission allowances market:

The first objective is to perform a comprehensive analysis of the key variables influencing the emission allowances market, such as allowance prices, emissions by sector and country, and macroeconomic indicators. This analysis will provide a deep understanding of the fundamental dynamics governing the EU ETS market, enabling the identification of main driving forces and challenges.

2. Analyze the correlation between economic and political variables and the emission allowances market:

The goal is to explore how economic and political variables, including regulatory decisions and macroeconomic conditions, affect the prices and quantities of emission allowances. Identifying these correlations is essential to understanding the market's sensitivity to different factors and how these can be manipulated or anticipated to achieve emission reduction goals.

3. Develop predictive models for significant market changes:

The third objective focuses on building predictive models to estimate the likelihood of significant market changes using historical variables. This includes the application of statistical and machine learning techniques to predict the direction and magnitude of market movements, which is crucial for strategic decision-making in emissions management.

3. Project definition

The project is structured into several fundamental parts, each addressing key aspects of the analysis and modeling of the EU carbon emissions market. The first part introduces the general context of the carbon market, describing its relevance and the main influencing factors. The second part focuses on the theoretical framework underpinning the analysis, including a review of the literature and the statistical and machine learning models employed. The third part details the process of data collection, selection, and preparation of the variables used in the study.

The fourth part is dedicated to developing the predictive models, describing the applied techniques, training and validation processes, and the obtained results. The fifth part offers a comparison of the implemented models, analyzing their predictive capabilities and result accuracy. Lastly, the sixth part includes the bibliography supporting and contextualizing the analyses conducted.

Particular emphasis is placed on the fourth part, which forms the core of the project. This section begins with a detailed description of the tools and methodologies used to construct the predictive

models, highlighting the differences between statistical and machine learning approaches. It then addresses the variable selection process through correlation analysis and dimensionality reduction techniques, ensuring the models focus on the most relevant variables. The development and implementation of the models are subsequently presented, analyzing their training and evaluation to ensure robustness and accuracy. Finally, the results are displayed and interpreted within the context of the carbon emissions market.

The analysis is based on a dataset that includes economic, energy, and financial information combined with carbon emission prices in the European Union. The modeling techniques range from traditional methods like linear regressions to advanced approaches like neural networks and long-term memory models. This comprehensive approach captures both linear and nonlinear relationships in the data, providing a thorough and precise view of the carbon emissions market dynamics.

4. Results

The study involved analyzing various predictive models applied to carbon price forecasting within the EU ETS market. These models included linear approaches, decision tree-based methods, and advanced neural networks. The main objective was to evaluate the accuracy and capability of each model in capturing the complexity and volatility of this market, which is shaped by multiple economic, energy, and regulatory factors.

Linear models, such as multiple and polynomial regression, exhibited significant limitations, reflected in their high prediction errors (RMSE) and negative determination coefficients (R^2). Although regularization techniques like Lasso and Ridge slightly improved performance, these approaches proved insufficient for modeling the data's complex nature. Meanwhile, tree-based methods, including Random Forest, Gradient Boosting, and XGBoost, provided a better fit, capturing nonlinear relationships with greater precision. However, these models also failed to adequately explain the observed variability, consistently showing negative R^2 values across configurations.

In contrast, neural networks, especially Long Short-Term Memory (LSTM) networks, excelled in the analysis. LSTM's ability to handle temporal relationships enabled significantly more accurate

predictions. In its optimized configuration, this model achieved an RMSE of 0.4408 and an R^2 of 0.9757, highlighting its superior capability to capture the dynamics of the carbon market. Although the optimized Multilayer Perceptron (MLP) model demonstrated significant improvements over linear and ensemble methods, it could not match the performance of LSTMs.

These results suggest that traditional approaches are limited in modeling a problem as complex as carbon price forecasting, whereas advanced neural networks offer a more precise and robust solution for this task.

5. Conclusions

The comparative analysis of predictive models identified strengths and limitations within the EU ETS market context, where volatility and the interplay of multiple factors necessitate sophisticated approaches. While linear and tree-based models provided an initial perspective on market dynamics, their performance was inadequate in effectively capturing the temporal and nonlinear complexities characteristic of this problem.

Conversely, the use of advanced neural networks, particularly LSTMs, emerged as a significantly superior strategy. These models not only markedly reduced prediction errors but also demonstrated a unique ability to integrate temporal dependencies and capture complex data fluctuations. This advancement positions LSTMs as the most suitable tool for such tasks and underscores the importance of employing methodologies tailored to the specific nature of the analyzed data.

Collectively, this study reaffirms that traditional solutions are limited when addressing the inherent complexity of the EU ETS market. Innovations in machine learning and neural networks can provide a more solid and effective predictive framework. These findings have implications not only for carbon price forecasting but also open new opportunities to tackle similar challenges in other complex and dynamic markets.

6. References

- [1] D. Ellerman, F. J. Convery, and C. de Perthuis. Pricing Carbon: The European Union Emissions Trading Scheme. Cambridge University Press, 2010.

- [2] Organisation for Economic Co-operation and Development (OECD). Carbon Pricing: What Role for Border Carbon Adjustments. Technical Report, 2021.
- [3] Journal of Cleaner Production. The Impact of Carbon Price and Firm-Level Emissions on Swedish Firms' Profits. 2023.
- [4] Energy Economics. The Environmental-Financial Performance Nexus of EU ETS Firms: A Quantile Regression Approach. 2022.
- [5] K. S. Rogge, M. Schneider, and V. H. Hoffmann. The Innovation Impact of the EU Emission Trading System: Findings of Company Case Studies in the German Power Sector. *Ecological Economics*, 70(2011), 513–523.
- [6] BI Norwegian Business School. EU ETS & the Maritime Shipping Industry: Barriers Compromising the Economic Efficiency and Effectiveness of the System. 2024.
- [7] Cambridge Press. The European Union's Emissions Trading System in Perspective. 2019.
- [8] R. Rafaty, G. Dolphin, and F. Pretis. Carbon Pricing and the Elasticity of CO₂ Emissions. Energy Policy Research Group, University of Cambridge, Working Paper 2035, 2020.
- [9] A. Wach. The European Union Emissions Trading Scheme: Failure Analysis and Assessment of Market Stability Reserve Solution. Norwegian School of Economics, Bergen, 2016.
- [10] European Commission. The EU Emissions Trading System. European Commission, 2021.
- [11] L. H. Goulder and A. R. Schein. Carbon Pricing in the EU: Evaluation of Different EU ETS Reform Options. Technical Report, 2022.
- [12] C. Flachsland, N. Koch, and S. Fuss. A Framework for Assessing the Performance of Cap-and-Trade Systems: Insights from the EU ETS. *Review of Environmental Economics and Policy*, 2018.
- [13] M. A. Mehling and R. A. Ritz. Emissions Trading Hybrids: The Case of the EU ETS. *Energy Economics*, 2022.
- [14] European Commission, Climate Action, EU ETS Emissions Cap. [En línea]. Disponible en: https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/eu-ets-emissions-cap_en.
- [15] M. Cames and M. Frerk. The Best (and Worst) of GHG Emission Trading Systems: Comparing the EU ETS with Its Followers. *Journal of Environmental Policy*, 2019.
- [16] R. Martin, M. Muûls, and U. J. Wagner. Allocation and Competitiveness in the EU Emissions Trading Scheme: Policy Overview. Carbon Markets Series, 2020.
- [17] F. Laurent. The European Ships' Monitoring, Reporting and Verification (MRV): Pre-evaluation of a Regional Regulation on Carbon Dioxide Inventory. IAME Conference, 2016.
- [18] J. Verschuuren and F. Fleurke. Report on the Legal Implementation of the EU ETS at Member State Level. Tilburg University, 2014.

- [19] J. Kruger and C. Egenhofer. Confidence through Compliance in Emissions Trading Markets. *Sustainable Development Law & Policy*, 2005.
- [20] T. Santier. Monitoring, Reporting, and Verification in a Carbon Border Adjustment Mechanism. 2023.
- [21] M. Zhang, Y. Liu, and Y. Su. Comparison of Carbon Emission Trading Schemes in the European Union and China. *Climate*, 2017.
- [22] P. Gailhofer and V. Graichen. Monitoring, Reporting and Verification in the EU ETS: Legal Design Considerations. 2024.
- [23] C. Warnecke. Can CDM Monitoring Requirements Be Reduced While Maintaining Environmental Integrity?. *Climate Policy*, 2014.
- [24] F. Deane, A. Huggins, and M. S. Karim. Measuring, Monitoring, Reporting and Verification of Shipping Emissions. *Review of European, Comparative & International Environmental Law*, 2019.
- [25] K. Kulovesi, E. Mekkonen, y T. Voigt. Enforcement of the EU ETS in the Member States. *Journal of Environmental Law*, 2021.
- [26] Red Nacional de Derechos de Emisión (RENADE), Taxes Renade. [En línea]. Disponible en: <https://www.renade.es/ing/Information/Taxes-Renade>.
- [27] J. Schleich, R. Betz, y K. Rogge. EU Emissions Trading – Better Job Second Time Around. *European Environment*, 2007.
- [28] J. Schleich, R. Betz, and K. Rogge. EU Emissions Trading – Better Job Second Time Around?. *European Environment*, 17(5), 279–294, 2007.
- [29] J. Wettestad and T. Jevnaker. *Rescuing EU Emissions Trading: The Climate Policy Flagship*. Springer, 2016.
- [30] J. B. Skjærseth and J. Wettestad. EU Emissions Trading: Initiation, Decision-Making and Implementation. *Global Environmental Politics*, 8(1), 117–141, 2008.
- [31] European Commission. *The EU Emissions Trading System (EU ETS)*. Technical Report, European Commission, 2021.
- [32] G. Perino y M. Willner. Procrastinating Reform: The Impact of the Market Stability Reserve on the EU ETS. *Journal of Environmental Economics and Management*, 2016.
- [33] C. Flachsland, R. Marschinski, and O. Edenhofer. To Link or Not to Link: Benefits and Disadvantages of Linking Cap-and-Trade Systems. *Climate Policy*, 9(4), 358–372, 2009.
- [34] J. Köhler, M. Grubb, and D. Popp. Transition Pathways and Risk Analysis for Climate Change Mitigation. *Climate Policy*, 19(3), 342–358, 2019.
- [35] M. Duwe, J. Graichen, y H. Böttcher. Can Current EU Climate Policy Reliably Achieve Climate Neutrality by 2050?. *Ecologic Institute*, 2023.

- [36] J. Boratyński, M. Pyrka, S. Lizak, I. Tobiasz, y R. Jeszke. Reform of the Market Stability Reserve (MSR) in the “Fit for 55” Package. IOS-PIB, 2022.
- [37] M. Wiels and J. C. Defraigne. The EU's Climate Ambitions: Assessment of the Social Cost of the Fit for 55 Package. Université Catholique de Louvain, 2023.
- [38] J. C. Richstein, É. J. L. Chappin, y L. J. de Vries. The Market (In-)Stability Reserve for EU Carbon Emission Trading: Why It Might Fail and How to Improve It. Utilities Policy, 2015.
- [39] D. Seibert, P. Kasten, J. Graichen, y N. Wissner. EU 2040 Climate Target: Contributions of the Transport Sector. Oeko Institute, 2024.
- [40] H. Rienks y B. Moore. EU Innovation Funding for Climate Neutrality. Wageningen University, 2023.
- [41] T. Hesse y S. Braungardt. EU 2040 Climate Target: Contributions of the Buildings Sector. Oeko Institute, 2024.
- [42] B. Chen and R. Wu. Comparative Analysis of the Carbon Market in the EU and China. Journal of Environmental Policy, 2023.
- [43] F. Annunziata. Emission Allowances as Financial Instruments. In Governance, Financial Stability and Financial Markets, Springer, 2021.
- [44] A. Michaelowa and I. Shishlov. Evolution of International Carbon Markets: Lessons for the Paris Agreement. Wiley Interdisciplinary Reviews: Climate Change, 10(6), e613, 2019.
- [45] M. Li and M. Duan. Exploring Linkage Opportunities for China's Emissions Trading System Under the Paris Targets: EU-China and Japan-Korea-China Cases. Energy Economics, 2021.
- [46] World Bank, Carbon Pricing Dashboard. [En línea]. Disponible en: <https://carbonpricingdashboard.worldbank.org/>.
- [47] C. Müller. Rethinking Voluntary Carbon Credit Markets and the EU ETS: A Critical Review. Fulda University of Applied Sciences, 2024.
- [48] D. Tänzler, J. Groß, and C. Warnecke. Interactions Between New Market Mechanisms and Emissions Trading Schemes: Opportunities and Prospects for Countries to Use Article 6 of the Paris Agreement. German Federal Ministry for the Environment, 2018.
- [49] O. M. Hladka, I. M. Karpovich, and A. Y. Sabaniuk. Artificial Neural Networks for Analysis of Stock Market Dynamics: Applications of Python and Jupyter Notebooks. Vernadsky Tech Journals, 2024.
- [50] E. Champa-Bujaico, A. M. Díez-Pascual, A. Lomas Redondo, and P. Garcia-Diaz. Optimization of mechanical properties of multiscale hybrid polymer nanocomposites: A combination of experimental and machine learning techniques. Composites Part B, vol. 269, p. 111099, 2024.

- [51] R. Bhattacharjee, K. Botchway, J. C. Pashin, G. Chakraborty, and P. Bikkina, "Machine learning-based prediction of CO₂ fugacity coefficients: Application to estimation of CO₂ solubility in aqueous brines as a function of pressure, temperature, and salinity," *International Journal of Greenhouse Gas Control*, vol. 128, p. 103971, 2023.
- [52] T. Souza, J. Herrera, A. Johnson, and K. B. Wright. Detection and classification of pneumonia using the Orange3 data mining tool. *International Journal of Computational Intelligence Systems*, vol. 14, no. 3, pp. 250-263, 2023.
- [53] F. Cheng, C. Yang, C. Zhou, L. Lan, H. Zhu, and Y. Li. Simultaneous Determination of Metal Ions in Zinc Sulfate Solution Using UV–Vis Spectrometry and SPSE-XGBoost Method. *Sensors*, vol. 20, no. 17, p. 4936, 2020.
- [54] K. Y. Chan, B. Abu-Salih, R. Qaddoura, A. M. Al-Zoubi, V. Palade, D.-S. Pham, J. Del Ser, and K. Muhammad. "Deep neural networks in the cloud: Review, applications, challenges and research directions." *Neurocomputing*, vol. 545, p. 126327, 2023.
- [55] B. Yang, Z. Guo, L. Wang, J. He, B. Xia, and S. Vakily. Updated Global Navigation Satellite System Observations and Attention-Based Convolutional Neural Network–Long Short-Term Memory Network Deep Learning Algorithms to Predict Landslide Spatiotemporal Displacement. *Remote Sens.*, vol. 15, no. 20, p. 4971, 2023.

Índice de la memoria

1. Sistema de Comercio de Emisiones de la UE (EU ETS)	31
1.1. Introducción al EU ETS	31
1.2. Funcionamiento del EU ETS	32
1.3. Monitoreo, Reporte y Verificación (MRV)	39
1.4. Desarrollo y Fases del EU ETS (2005-2020)	42
1.5. Marco Legislativo y Reformas del EU ETS	45
1.6. Impacto y Objetivos Climáticos del EU ETS	47
1.7. Integridad y Supervisión del Mercado de Carbono	49
1.7.1. Supervisión y Prevención de Abusos	49
1.8. Mercado Internacional de Carbono y Cooperación	50
1.9. Diplomacia en Precios Internacionales de Carbono	52
2. Objetivos del trabajo	53
2.1. Objetivo General	54
2.2. Objetivos Específicos	54
2.3. Relevancia de los Objetivos	55
2.4. Preguntas Clave del Trabajo	55
3. Herramientas y Técnicas de Programación para el Análisis y Modelado de Datos	56
3.1. Python	56
3.2. Jupyter Notebooks	58
4. Estudio de las Variables	60
4.1. Presentación y Clasificación de las Variables	61
4.2. Análisis de Correlaciones y Selección de Variables	66
4.3. Análisis Descriptivo de las Variables	80
5. Análisis de Clustering	97
5.1. Propósito y Contexto del Clustering	97
5.2. Desarrollo y Resultados del Análisis	98

6.	Modelos de Predicción	103
6.1.	Regresión Lineal Múltiple.....	105
6.2.	Regresión Polinómica	109
6.3.	Lasso y Ridge	114
6.4.	Árbol de Decisión.....	123
6.5.	Random Forest.....	130
6.6.	Gradient Boosting.....	137
6.7.	Extreme Gradient Boosting	145
6.8.	Multilayer Perceptron (MLP).....	152
6.9.	Long Short-Term Memory (LSTM)	159
6.10.	Comparativa de los Modelos.....	167
7.	Conclusión.....	174
8.	Bibliografía	176

Índice de figuras

Figura 1: Estructura del límite y asignación de permisos del EU ETS en la Fase 4 tras la revisión de 2023 [14].....	34
Figura 2: Instrumentos de Fijación de Precios de Carbono Implementados a Nivel Global en 2024 [46].....	52
Figura 3: Mapa de calor de las variables precios de electricidad.....	68
Figura 4: Explicación acumulada de la varianza (PCA).....	69
Figura 5: Matriz de correlación de Pearson.....	71
Figura 6: Matriz de correlación de Spearman.....	73
Figura 7: Matriz de correlación de Kendall.....	75
Figura 8: Gráfico de serie temporal de PIB.....	82
Figura 9: Gráfico de serie temporal de la producción de acero.....	83
Figura 10: Histograma del precio del petróleo.....	85
Figura 11: Boxplot del precio del petróleo.....	86
Figura 12: Gráfico de tendencia temporal del precio del petróleo.....	87
Figura 13: Histograma del precio del gas natural.....	88
Figura 14: Boxplot del precio del gas natural.....	89
Figura 15: Gráfico de tendencia temporal del precio del gas natural.....	90
Figura 16: Histograma de MSCI Europe.....	92
Figura 17: Diagrama de caja de MSCI Europe.....	93
Figura 18: Gráfico de serie temporal de MSCI Europe.....	94
Figura 19: Histograma del tipo de cambio USD/EUR.....	96
Figura 20: Gráfico de serie temporal del tipo de cambio USD/EUR.....	97
Figura 21: Método del codo.....	99
Figura 22: Clusters visualizados en 2D.....	101
Figura 23: Valores promedio de las variables en cada cluster.....	101
Figura 24: Relación entre los residuos y las predicciones del modelo de regresión lineal múltiple.....	107
Figura 25: Histograma de la distribución de los residuos del modelo de regresión lineal múltiple.....	107
Figura 26: Comparación entre los valores reales y las predicciones del modelo de regresión lineal múltiple.....	108
Figura 27: Comparación temporal de predicciones y valores reales del modelo de regresión lineal múltiple.....	108
Figura 28: Evaluación del rendimiento de la regresión polinómica según el grado del polinomio.....	110

Figura 29: Relación entre los residuos y las predicciones del modelo de regresión polinómica	112
Figura 30: Histograma de la distribución de los residuos del modelo de regresión polinómica	112
Figura 31: Comparación entre los valores reales y las predicciones del modelo de regresión polinómica.....	113
Figura 32: Comparación temporal entre los valores reales y las predicciones del modelo de regresión polinómica (grado 6)	113
Figura 33: Comparación temporal entre los valores reales y las predicciones del modelo Lasso optimizado	117
Figura 34: Relación entre los residuos y las predicciones del modelo Lasso optimizado	118
Figura 35: Histograma de la distribución de los residuos del modelo Lasso optimizado.....	118
Figura 36: Comparación entre los valores reales y las predicciones del modelo Lasso optimizado	119
Figura 37: Comparación temporal entre los valores reales y las predicciones del modelo Ridge optimizado	121
Figura 38: Relación entre los residuos y las predicciones del modelo Ridge optimizado.....	121
Figura 39: Histograma de la distribución de los residuos del modelo Ridge optimizado.....	122
Figura 40: Comparación entre los valores reales y las predicciones del modelo Ridge optimizado	122
Figura 41: Estructura de un Árbol de Decisión [50].....	125
Figura 42: Visualización del Árbol de Decisión optimizado.....	127
Figura 43: Relación entre los residuos y las predicciones de los modelos Árbol de Decisión	128
Figura 44: Histograma de la distribución de los residuos de los modelos Árbol de Decisión	128
Figura 45: Comparación entre los valores reales y las predicciones de los modelos Árbol de Decisión	129
Figura 46: Comparación temporal entre los valores reales y las predicciones del modelo de Árbol de Decisión optimizado	129
Figura 47: Estructura del modelo Random Forest [51].....	132
Figura 48: Visualización de un árbol individual del modelo Random Forest optimizado	134
Figura 49: Relación entre los residuos y las predicciones de los modelos Random Forest	134
Figura 50: Histograma de la distribución de los residuos de los modelos Random Forest	135
Figura 51: Comparación entre los valores reales y las predicciones de los modelos Random Forest	136
Figura 52: Comparación temporal entre los valores reales y las predicciones del modelo Random Forest optimizado.....	136
Figura 53: Proceso secuencial del algoritmo Gradient Boosting [52].....	139

Figura 54: Comparación temporal entre los valores reales y las predicciones del modelo Gradient Boosting optimizado.....	141
Figura 55: Visualización del árbol de decisión optimizado del modelo Gradient Boosting	142
Figura 56: Relación entre los residuos y las predicciones de los modelos Gradient Boosting.....	142
Figura 57: Distribución de los residuos de los modelos Gradient Boosting base y optimizado	143
Figura 58: Comparación entre los valores reales y las predicciones de los modelos Gradient Boosting.....	144
Figura 59: Estructura del algoritmo XGBoost: construcción secuencial y optimización [53]	148
Figura 60: Comparación temporal entre los valores reales y las predicciones del modelo XGBoost optimizado	148
Figura 61: Relación entre los residuos y las predicciones de los modelos XGBoost	149
Figura 62: Histograma de la distribución de los residuos de los modelos XGBoost	150
Figura 63: Comparación entre los valores reales y las predicciones de los modelos XGBoost	151
Figura 64: Estructura de una Red Neuronal Multicapa (Multilayer Perceptron) [54]	153
Figura 65: Desempeño del modelo MLP no optimizado: predicciones vs valores reales.....	155
Figura 66: Progreso del Entrenamiento del modelo MLP no optimizado	155
Figura 67: Desempeño del modelo MLP optimizado: predicciones vs valores reales.....	157
Figura 68: Progreso del Entrenamiento del modelo MLP optimizado.....	158
Figura 69: Predicción de los modelos MLP vs valores reales.....	159
Figura 70: Estructura de una red LSTM [55]	162
Figura 71: Desempeño del modelo LSTM: predicciones vs valores reales	163
Figura 72: Progreso del entrenamiento del modelo LSTM.....	164
Figura 73: Desempeño del modelo LSTM optimizado: predicciones vs valores reales	165
Figura 74: Progreso del entrenamiento del modelo LSTM optimizado	166
Figura 75: Predicción de modelos LSTM vs valores reales.....	166
Figura 76: Comparación de todos los modelos: predicción vs. valores reales	172

Índice de tablas

Tabla 1: Participación de Permisos Generales (Capítulo III de la Directiva 2003/87/EC).....	36
Tabla 2: Permisos de Aviación (Capítulo II de la Directiva 2003/87/EC).....	37
Tabla 3: Tarifas de la Sección Española del Union Registry para Cuentas de Emisiones de Carbono [26].....	41
Tabla 4: Valores VIF de la selección inicial de variables	77
Tabla 5: Valores VIF de la selección final de variables	78
Tabla 6: Correlaciones de las variables seleccionadas	79
Tabla 7: Correlaciones de las variables eliminadas	80
Tabla 8: Comparación entre Lasso y Ridge.....	115
Tabla 9: Tabla comparativa de los resultados de los modelos	174

Índice de outputs

Output 1: Análisis descriptivo de PIB	81
Output 2: Análisis descriptivo de la producción de acero	83
Output 3: Análisis descriptivo del precio del petróleo	84
Output 4: Análisis descriptivo del precio del gas natural.....	88
Output 5: Análisis descriptivo de MSCI Europe.....	91
Output 6: Análisis descriptivo del tipo de cambio USD/EUR.....	95
Output 7: Métricas modelo de regresión lineal múltiple	106
Output 8: Métricas modelo de regresión polinómica (grado 6)	111
Output 9: Métricas modelo Lasso.....	116
Output 10: Métricas modelo Lasso optimizado	116
Output 11: Métricas del modelo Ridge	120
Output 12: Métricas del modelo Ridge optimizado	120
Output 13: Métricas modelo Árbol de Decisión.....	126
Output 14: Métricas modelo Árbol de Decisión optimizado.....	126
Output 15: Métricas modelo Random Forest.....	132
Output 16: Métricas modelo Random Forest optimizado.....	133
Output 17: Métricas modelo Gradient Boosting.....	140
Output 18: Métricas modelo Gradient Boosting optimizado	141
Output 19: Métricas modelo Extreme Gradient Boosting	147
Output 20: Métricas modelo Extreme Gradient Boosting optimizado	148

1. Sistema de Comercio de Emisiones de la UE (EU ETS)

1.1. Introducción al EU ETS

1.1.1. Objetivo y alcance

El Sistema de Comercio de Emisiones de la Unión Europea (EU ETS) fue creado en 2005 y se ha convertido en el pilar de la estrategia de la Unión Europea para reducir emisiones de gases de efecto invernadero (GEI), específicamente en sectores altamente contaminantes como la generación de energía, la industria pesada y la aviación [1]. En 2024, el alcance del EU ETS se expandirá para incluir también el transporte marítimo [2]. Este sistema abarca actualmente más de 10,000 instalaciones en los Estados Miembros de la UE y en países asociados como Islandia, Liechtenstein y Noruega, además del Reino Unido en lo referente a Irlanda del Norte [3]. La cobertura geográfica del EU ETS es amplia y complementa los esfuerzos globales en la lucha contra el cambio climático, siendo un modelo replicado por otras economías [4].

El EU ETS tiene el objetivo de proporcionar una solución económica y efectiva para limitar las emisiones en los sectores de más alto consumo energético. Su implementación permite a la UE avanzar de manera significativa hacia sus metas climáticas establecidas en la Ley Climática Europea, la cual busca alcanzar la neutralidad en carbono para 2050 y reducir un 55% de las emisiones para 2030 en comparación con los niveles de 1990 [5]. Esta herramienta responde a la necesidad de incentivar económicamente la reducción de emisiones, trasladando los costes del daño ambiental a los agentes responsables, lo cual resulta en un modelo económico sostenible y proactivo en la transición hacia energías renovables y prácticas industriales limpias [6].

1.1.2. Principios fundamentales (“cap and trade”)

El EU ETS funciona bajo el sistema de “límite y comercio” (“cap and trade”), un enfoque de control de emisiones basado en el establecimiento de un límite anual o “cap” de emisiones. Este límite se traduce en permisos de emisión, que otorgan a cada operador el derecho a emitir una tonelada de CO₂ equivalente (tCO₂eq) [7]. A través de este sistema, el EU ETS establece un

mercado en el que las empresas pueden comprar y vender estos permisos, incentivando la reducción de emisiones mediante la introducción de un precio al carbono [8].

Cada año, el límite se reduce de manera progresiva, lo que obliga a las empresas a optimizar su uso de permisos o adoptar tecnologías más limpias para evitar un aumento en sus costes operativos [4]. Esto promueve la transición hacia prácticas de producción más sostenibles y eficientes en términos de carbono, ya que las empresas con menores emisiones pueden vender sus permisos sobrantes a aquellas que aún no han logrado reducir sus niveles de emisión [2]. Este modelo flexible permite a las empresas ajustar su estrategia de reducción de emisiones en función de su estructura y capacidad, sin imponer restricciones inflexibles que puedan obstaculizar su competitividad en el mercado [6].

Además, este sistema incluye mecanismos de control y transparencia como el Union Registry, que registra todas las transacciones y permisos en circulación, asegurando que cada operación cumpla con los estándares de monitoreo, reporte y verificación establecidos en el marco regulador del EU ETS. El precio de los permisos se establece en el mercado de carbono de la UE y es regulado en función de la oferta y demanda, lo que hace que el costo de emitir carbono sea más alto con el tiempo, alineándose con la reducción gradual del límite de emisiones [1]. Desde su creación, el EU ETS ha demostrado ser eficaz en la creación de un mercado estable para el carbono, generando una reducción del 47% en las emisiones de los sectores cubiertos entre 2005 y 2023, al mismo tiempo que ha recaudado miles de millones de euros para financiar la transición energética en Europa [4].

1.2. Funcionamiento del EU ETS

1.2.1. Límite de emisiones y permisos

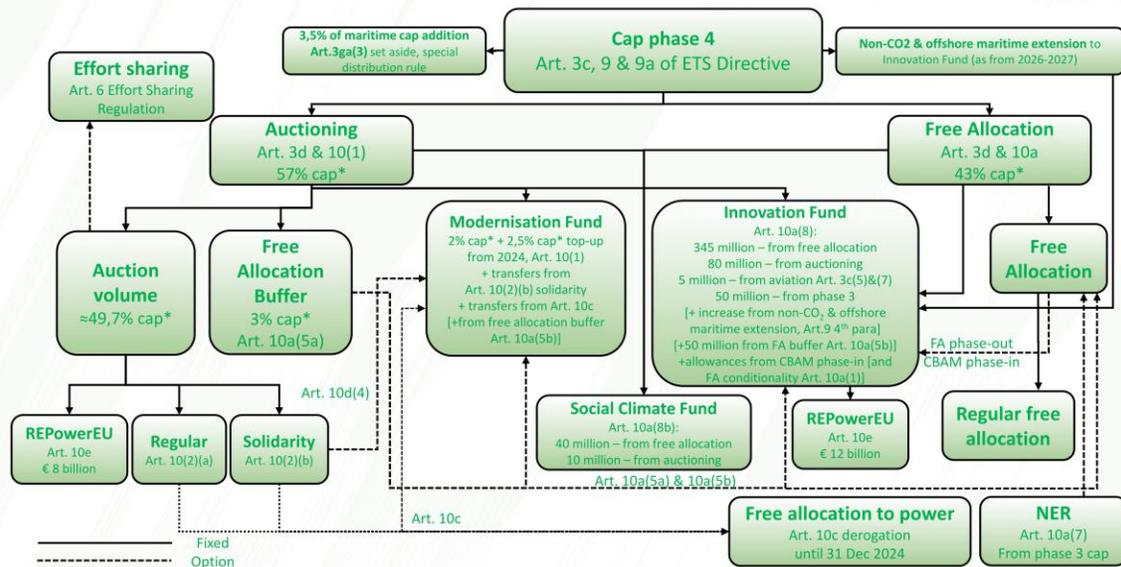
El EU ETS establece un límite máximo de emisiones de gases de efecto invernadero (GEI) permitido para cada periodo. Este sistema no solo define el volumen máximo de emisiones permitido, sino que establece la base para la disminución progresiva de emisiones en sectores clave [9]. Desde su implementación, el límite se ha ajustado de forma continua, con el objetivo de reducir el volumen de permisos disponibles y aumentar los costos de emisión de carbono. Con el tiempo,

el sistema ha transitado hacia un límite unificado para toda la UE, lo que ha contribuido a mejorar la eficiencia y efectividad del sistema en la reducción de emisiones [10].

A partir de 2013, se adoptó un límite común para todos los sectores y países participantes, sustituyendo los límites nacionales previos, lo cual fortaleció la coherencia del sistema en toda la región [11]. Este límite decrece anualmente mediante un factor de reducción, el cual se incrementó en 2021 al 2.2% anual y, tras la revisión de 2023, subirá al 4.3% entre 2024 y 2027, alcanzando el 4.4% a partir de 2028 [12]. Estas disminuciones son necesarias para asegurar que el sistema EU ETS contribuya a alcanzar la meta climática de una reducción del 62% en las emisiones de GEI para 2030 respecto a los niveles de 2005. Además de este ajuste continuo, se planifican rebajas específicas del límite en 2024 y 2026, que reducirán el volumen total de permisos disponibles, lo que obligará a las industrias a mejorar sus prácticas para adaptarse al nuevo entorno de mercado de bajas emisiones [13].

La siguiente imagen ilustra la estructura detallada del límite de emisiones y la distribución de permisos en la Fase 4 del EU ETS, implementada tras la revisión de 2023. Muestra cómo se asignan los permisos mediante subastas y asignaciones gratuitas, así como la distribución de fondos específicos, incluyendo el Fondo de Modernización, el Fondo de Innovación y el Fondo Social Climático. Estos mecanismos garantizan una gestión equilibrada de los permisos en el mercado de carbono y apoyan la transición hacia una economía de bajas emisiones.

Phase 4 EU ETS cap for general allowances (post 2023 revision)



* The percentages for the share of various elements in the cap are calculated in relation to the part of the Art. 9 cap that is subject to Art. 10 and 10a distribution, i.e. after deducting the 3,5% of the maritime cap referred to in Art. 3ga(3) to be attributed to EL, CY and MT considering their ratio of shipping companies based on data available for the period from 2018 to 2020, and the share of the cap from the non-CO2 and offshore maritime extension.
 * These percentages do not take into account the Art. 3c cap, where free allocation will be phased out by 2026, and a review will address scope for 2027-30, Art 28b(3).

Figura 1: Estructura del límite y asignación de permisos del EU ETS en la Fase 4 tras la revisión de 2023 [14]

1.2.2. Asignación de permisos y subastas

La asignación de permisos se realiza principalmente mediante subastas, basándose en el principio de “quien contamina, paga” [15]. Este enfoque asegura que las empresas internalicen el coste de sus emisiones, creando un incentivo directo para que inviertan en tecnologías y prácticas más limpias. Desde 2013, la subasta se ha consolidado como el método predeterminado de asignación de permisos para la mayoría de los sectores, especialmente el de generación de energía, aunque algunos sectores industriales que enfrentan el riesgo de “fuga de carbono” reciben permisos gratuitos para asegurar su competitividad global y evitar que se trasladen fuera de la UE [16]. Los permisos gratuitos se asignan mediante un sistema de benchmarking, que compara la eficiencia de los sectores según la intensidad de emisiones del 10% más eficiente, impulsando a las empresas a reducir sus emisiones para mantener una ventaja competitiva.

Para asegurar la equidad y fomentar la solidaridad entre los Estados Miembros, el 90% de los permisos subastados se distribuyen proporcionalmente entre los países según sus emisiones históricas, mientras que el 10% restante se destina a Estados Miembros con economías en transición.

La siguiente tabla muestra la asignación porcentual de permisos de emisión generales que cada Estado Miembro de la UE puede subastar entre 2021 y 2030, en conformidad con el Capítulo III de la Directiva 2003/87/EC. Estos permisos están destinados a instalaciones industriales y generadoras de energía, y su distribución se basa en las emisiones históricas verificadas de cada país, reflejando así su participación en las emisiones totales reguladas bajo el EU ETS.

State	Auction share (%)
Austria	1,549511512
Belgium	2,570508102
Bulgaria	2,683627316
Croatia	0,519852619
Cyprus	0,287177271
Czech Republic	5,021898139
Denmark	1,394176362
Estonia	0,853364580
Finland	1,861257204
France	6,094561928
Germany	22,290594909
Greece	3,815882431
Hungary	1,526557204
Iceland	0,043450983
Ireland	1,041937115
Italy	10,521550682
Latvia	0,200953534
Liechtenstein	0,000972381
Lithuania	0,433065735
Luxemburg	0,122010786
Malta	0,111715994
Netherlands	3,730700676
Norway	0,865251836
Poland	13,000319671

Portugal	1,933850872
Romania	4,767806687
Slovakia	1,602124134
Slovenia	0,485701471
Spain	9,519452381
Sweden	0,904220479
United Kingdom – Northern Ireland	0,245945006

Tabla 1: Participación de Permisos Generales (Capítulo III de la Directiva 2003/87/EC)

A continuación, se presenta la asignación de permisos de emisión específicos para el sector de la aviación, distribuidos entre los Estados Miembros de acuerdo con las emisiones verificadas del sector en 2018 para vuelos dentro del Espacio Económico Europeo, según lo dispuesto en el Capítulo II de la Directiva 2003/87/EC. Esta tabla especifica los porcentajes de permisos que cada país puede subastar para vuelos de aviación civil bajo el EU ETS.

State	Auction share (%)
Austria	2,124352274
Belgium	2,884043556
Bulgaria	0,962095516
Croatia	0,770787777
Cyprus	0,698769072
Czech Republic	1,021630732
Denmark	2,589242238
Estonia	0,269340406
Finland	2,141104211
France	11,28357633
Germany	15,85911782
Greece	5,222468391
Hungary	0,97207424
Iceland	0,912691877

Ireland	1,928162337
Italy	12,29799421
Latvia	0,542686113
Lithuania	0,441761613
Luxemburg	0,354887922
Malta	0,599063928
Netherlands	4,288464549
Norway	4,219779278
Poland	2,95885866
Portugal	3,885085782
Romania	1,861835804
Slovakia	0,15910511
Slovenia	0,128561379
Spain	14,8981547
Sweden	3,724304171

Tabla 2: Permisos de Aviación (Capítulo II de la Directiva 2003/87/EC)

Los ingresos generados por las subastas se asignan a proyectos que promuevan la transición verde a nivel nacional, incluyendo mejoras en eficiencia energética, fomento de energías renovables y desarrollo de transporte sostenible. Entre 2013 y 2023, el EU ETS generó más de 200 mil millones de euros en ingresos, financiando programas como el Fondo de Innovación, el Fondo de Modernización y, más recientemente, el Fondo Social Climático, que apoya a las poblaciones más vulnerables frente a los impactos del costo del carbono.

1.2.3. Reserva de estabilidad del mercado (MSR)

La Reserva de Estabilidad del Mercado (MSR) es un mecanismo clave introducido en 2018 para enfrentar la volatilidad de los precios en el mercado de carbono. Tras la crisis económica de 2008 y el exceso de permisos que surgió, el precio del carbono disminuyó considerablemente, lo cual

redujo la efectividad del sistema como herramienta de reducción de emisiones. La MSR, operativa desde 2019, ajusta la oferta de permisos en el mercado al monitorear el número total de permisos en circulación (Total Number of Allowances in Circulation, TNAC). Cuando el TNAC supera los 833 millones, la MSR retira un porcentaje de permisos del mercado; cuando está por debajo de 400 millones, se liberan permisos adicionales para equilibrar la oferta y la demanda.

Desde 2023, la MSR también elimina permanentemente los permisos en exceso que superen un umbral específico, consolidando así el valor de los permisos restantes y asegurando que el sistema EU ETS mantenga su eficacia en la reducción de emisiones. Esta regulación de la oferta previene fluctuaciones extremas y proporciona estabilidad a corto plazo, lo cual es esencial para que las empresas puedan planificar sus inversiones en tecnologías de bajas emisiones de forma más confiable.

1.2.4. Uso de créditos internacionales

El EU ETS ha permitido el uso de créditos internacionales para facilitar el cumplimiento de las metas de emisiones a través de mecanismos del Protocolo de Kioto, como el Mecanismo de Desarrollo Limpio (CDM) y la Implementación Conjunta (JI). Estos créditos representan una tonelada de CO₂ evitada o retirada de la atmósfera y permiten a las empresas cumplir una parte de sus obligaciones comprando créditos generados por proyectos de reducción de emisiones en países en desarrollo. En la Fase 3 (2013-2020), el uso de estos créditos fue limitado mediante restricciones cualitativas y cuantitativas para asegurar la integridad del sistema, permitiendo solo aquellos créditos provenientes de proyectos específicos y de sectores que cumplan con altos estándares de transparencia y sostenibilidad.

Con la implementación del Acuerdo de París, la UE se ha orientado hacia un objetivo de reducción de emisiones principalmente interno, y desde 2021 ha restringido significativamente el uso de créditos internacionales en el EU ETS. Sin embargo, el artículo 6 del Acuerdo de París establece la posibilidad de futuros mecanismos de colaboración para la reducción de emisiones que podrían reintegrar los mercados de carbono en una plataforma global. Esto abriría nuevas oportunidades para que los países participen en el comercio de carbono de manera transparente y eficiente, manteniendo altos estándares ambientales y asegurando que las reducciones de emisiones

realizadas en cualquier parte del mundo contribuyan al compromiso de reducir los GEI a nivel global.

1.3. Monitoreo, Reporte y Verificación (MRV)

1.3.1. Ciclo de cumplimiento anual

El proceso de Monitoreo, Reporte y Verificación (MRV) es la columna vertebral del cumplimiento en el EU ETS, asegurando la precisión y transparencia de los datos de emisiones en todo el sistema [17]. Cada instalación y operador aéreo regulado debe contar con un plan de monitoreo aprobado, el cual establece los procedimientos para medir las emisiones de gases de efecto invernadero. Este plan, que es un requisito para obtener la licencia de operación, incluye metodologías detalladas y estándares específicos de medición que aseguran la consistencia de los datos recopilados en todos los Estados Miembros [18]. La implementación de estos planes permite que el sistema mantenga un alto nivel de precisión en sus reportes, lo que es esencial para el cumplimiento de las metas de reducción de emisiones establecidas por la UE y los países asociados [19].

Cada año, los operadores deben preparar un informe detallado de sus emisiones, abarcando el año anterior. Este informe debe ser verificado por una entidad acreditada antes del 31 de marzo del año siguiente, quien asegura que los datos cumplen con los estándares requeridos y reflejen fielmente las emisiones de la instalación [20]. Este proceso de verificación, obligatorio y riguroso, es fundamental para mantener la integridad del sistema [21]. Los verificadores no solo revisan la metodología empleada para medir las emisiones, sino también la calidad de los datos reportados, garantizando que cada etapa del proceso esté alineada con los estándares exigidos.

Posteriormente, antes del 30 de abril, los operadores deben rendir una cantidad de permisos de emisión que corresponda exactamente a sus emisiones verificadas, concluyendo así el ciclo anual de cumplimiento [22]. Este proceso también es auditado en todas sus fases, maximizando la confiabilidad de los datos y asegurando que las instalaciones cumplan con sus obligaciones. El enfoque en la verificación detallada fortalece la credibilidad del EU ETS, ya que las empresas que participan en el sistema saben que sus reportes estarán sujetos a un control exhaustivo, eliminando cualquier posible irregularidad en la contabilización de emisiones.

1.3.2. Union Registry y European Union Transaction Log (EUTL)

El Union Registry centraliza la supervisión de permisos, eliminando la necesidad de registros nacionales individuales y permitiendo un seguimiento eficiente de cada permiso de emisión [23]. Este registro digital anota todas las transacciones y mantiene actualizadas las asignaciones de permisos para cada instalación, funcionando de manera similar a un banco de datos financiero que documenta cada transferencia, emisión y verificación de permisos [24]. En complemento, el European Union Transaction Log (EUTL) supervisa de manera automática todas las transacciones entre cuentas, garantizando que cumplan con las normativas del EU ETS [25]. La infraestructura digital del Union Registry y el EUTL fortalece la trazabilidad de las transacciones y evita cualquier práctica de comercio ilícito en el sistema.

1.3.3. Costes y tasas administrativas

La siguiente tabla presenta las tarifas aplicables en la sección española del Registro de la Unión para diferentes tipos de cuentas en el EU ETS, incluyendo cuentas de operadores, cuentas de operadores aéreos, cuentas de negociación y cuentas de personas. Estas tarifas cubren los costos de apertura y mantenimiento de cuentas, así como los cargos asociados a las emisiones verificadas [23]. Los valores reflejan la estructura de costos diseñada para asegurar una gestión transparente y eficiente del sistema de permisos de emisión en el mercado de carbono [22].

FEES OF THE SPANISH SECTION OF THE UNION REGISTRY*

Operator Holding account, aircraft operator, maritime operator or regulated entity			
Account Opening	At the time of the opening		575 €
Verified emissions	Annual***	0,0045 € per ton of CO ₂ entered in the account up to 13.500	€ maximum per account and year
	Annual**		
Maintenance	For the purposes of determining verified emissions, the volume of verified emissions corresponding to the year prior to the accrual period will be taken as a reference. In the event that the opening occurs in a year in which there is no obligation to register verified emissions corresponding to the previous year, the amount of 115 euros will be applied.	Verified emissions ≥ 25.000 tCO _{2(e)}	170 €
		Verified emissions < 25.000 tCO _{2(e)}	115 €
TRADING ACCOUNT			
Trading Account Opening	At the time of the opening		805 €
Trading Account Maintenance	Annual**		1.150 €
NATIONAL KIOTO HOLDING ACCOUNT			
Person Holding Account Opening	At the time of the opening		805 €
Person Holding Account Maintenance	Annual**		285 €
OTHER FEES			
Certificate Issuance			60 €
Certificate Issuance (English and Spanish)			100 €

* With the exception of maritime sector, that the fee apply from 2024, for the rest of the account types, these fees will come into force in 2025.

**The opening year, it will be billed jointly with the account opening fee.

*** Will not be subject to the charge for this operation if the verified Emissions are less than 25.000 tCO_{2(e)}

Tabla 3: Tarifas de la Sección Española del Union Registry para Cuentas de Emisiones de Carbono [26]

1.3.4. Penalización

El EU ETS aplica sanciones estrictas a aquellos operadores que no cumplen con la entrega de permisos en el plazo establecido. En caso de incumplimiento, los operadores deben pagar una multa de 100 euros por tonelada de emisiones no cubierta por un permiso, y esta sanción se incrementa cada año en función del índice de precios al consumo de la UE [17]. Además, deben adquirir y entregar los permisos faltantes para regularizar su situación. Este sistema de sanciones promueve el cumplimiento puntual y crea una presión financiera considerable sobre aquellos que intentan evadir sus obligaciones de emisión [19]. La publicación de los nombres de las empresas que incumplen añade una capa de presión social, promoviendo una mayor responsabilidad y transparencia en el sistema.

1.3.5. Normativas de Monitoreo y Verificación

El cumplimiento de los requisitos de MRV está regulado mediante dos normativas específicas: el Reglamento de Monitoreo y Reporte (MRR) y el Reglamento de Acreditación y Verificación (AVR) [25]. El MRR define las metodologías y directrices para el monitoreo y reporte de emisiones, mientras que el AVR regula las condiciones de acreditación y las obligaciones de los verificadores independientes. Estas normativas proporcionan un marco unificado para la medición y verificación de emisiones, asegurando que los datos reportados sean consistentes y comparables en todos los países que participan en el EU ETS [21].

Para facilitar la implementación de estos reglamentos, la Comisión Europea ha desarrollado herramientas y plantillas que guían tanto a los operadores como a las autoridades en el proceso de monitoreo y reporte [20]. Esta asistencia incluye ejemplos de planes de monitoreo, informes de emisión, documentos de verificación y guías para mejorar la precisión y eficiencia en el reporte de emisiones. Además, sectores específicos como la aviación cuentan con herramientas adicionales y simplificadas, como el Eurocontrol's Simplified Monitoring Tool, que permite a pequeños emisores calcular y verificar sus emisiones de manera eficiente. Este enfoque armonizado asegura que todos los participantes del EU ETS, independientemente de su tamaño o sector, cumplan con las normativas de MRV de forma coherente y confiable, promoviendo así la transparencia y efectividad del sistema en su conjunto.

1.4. Desarrollo y Fases del EU ETS (2005-2020)

1.4.1. Fase 1 (2005-2007): Período de aprendizaje y establecimiento de precios de carbono

La primera fase del EU ETS fue diseñada como un período de “aprendizaje práctico” en el que la prioridad fue establecer las bases del sistema y permitir a los participantes comprender sus mecanismos. Durante estos tres años, el EU ETS se centró exclusivamente en las emisiones de dióxido de carbono (CO₂) provenientes de sectores intensivos en energía, principalmente generadores de electricidad e industrias pesadas [27]. Prácticamente la totalidad de los permisos se distribuyeron gratuitamente a las empresas en esta fase inicial, lo cual facilitó la participación

y permitió a las industrias adaptarse al concepto de comercio de emisiones sin incurrir en costos inmediatos.

Uno de los principales logros de esta fase fue la creación de un precio inicial para el carbono en Europa. Aunque no existían datos previos sólidos sobre las emisiones reales de los sectores incluidos, el sistema permitió establecer una infraestructura operativa y monitorear la actividad de los sectores regulados [28]. Sin embargo, al haberse basado en estimaciones, el número de permisos asignados superó ampliamente las emisiones reales, lo que generó un exceso de oferta en el mercado y llevó a que el precio de los permisos cayera a cero al final de esta fase, ya que no podían ser transferidos a la siguiente. Este desenlace puso en evidencia la necesidad de mejorar la precisión en la asignación de permisos y de incluir medidas de ajuste [1].

1.4.2. Fase 2 (2008-2012): Expansión geográfica y aumento de sectores regulados

La segunda fase coincidió con el primer período de compromiso del Protocolo de Kioto y supuso una expansión significativa del EU ETS tanto en términos de cobertura geográfica como de sectores [29]. Durante este periodo, Islandia, Liechtenstein y Noruega se unieron al sistema, extendiendo el ámbito del EU ETS más allá de las fronteras de la UE. Además, varios países comenzaron a incluir las emisiones de óxido nitroso (N₂O) derivadas de la producción de ácido nítrico, lo cual amplió la gama de GEI cubiertos por el sistema y permitió una mejor alineación con los compromisos de Kioto.

Para mejorar la eficacia del sistema y reducir el exceso de oferta de permisos, el límite de emisión se redujo aproximadamente un 6.5% respecto a 2005. La proporción de permisos gratuitos disminuyó ligeramente, pasando a aproximadamente el 90%, y varios Estados miembros comenzaron a introducir subastas de permisos para facilitar la creación de un mercado más dinámico. La penalización por no cumplir con los requisitos de entrega de permisos también se incrementó a 100 euros por tonelada, lo cual incentivó un cumplimiento más riguroso [30]. Durante esta fase, las empresas pudieron cumplir parcialmente sus obligaciones comprando créditos internacionales generados por proyectos de reducción de emisiones en países en desarrollo, alcanzando un total de 1.4 mil millones de toneladas de CO₂ equivalente. Esta iniciativa ayudó a

conectar el EU ETS con los mecanismos internacionales y fomentó el apoyo a proyectos de reducción de emisiones en países menos desarrollados.

Un cambio significativo en la Fase 2 fue la introducción de un sistema de registro centralizado: el Registro de la Unión reemplazó los registros nacionales y el Community Independent Transaction Log (CITL) fue sustituido por el European Union Transaction Log (EUTL), lo que permitió una mayor transparencia y trazabilidad en las transacciones [31]. También, en 2012 se incluyó la aviación dentro del EU ETS, aunque se limitó a vuelos internos al EEE y a vuelos con destino o procedencia de Suiza y el Reino Unido, marcando un primer paso hacia la regulación de las emisiones de este sector.

1.4.3. Fase 3 (2013-2020): Introducción de un límite a nivel de la UE y cambios en la asignación de permisos

La tercera fase del EU ETS introdujo reformas estructurales significativas y marcó un cambio hacia un sistema más armonizado y centralizado. En lugar de límites de emisión específicos por país, se implementó un límite único a nivel de la UE que cubría a todas las instalaciones de generación de electricidad y a la mayoría de las industrias intensivas en energía [32]. Este límite se redujo cada año en función de un factor de reducción lineal, lo cual aseguró que las emisiones disminuyeran de manera continua en toda la UE.

Uno de los cambios más importantes en esta fase fue el uso de la subasta como método principal para asignar permisos, estableciendo el principio de “quien contamina, paga”. Solo ciertos sectores, que podrían enfrentarse al riesgo de fuga de carbono, continuaron recibiendo una asignación de permisos gratuita bajo nuevas reglas armonizadas de asignación basadas en benchmarks de eficiencia. Se incluyeron también más sectores y gases bajo la regulación del EU ETS, ampliando así su alcance y adaptándolo a una mayor diversidad de actividades industriales [33]. Para estimular la innovación y la adopción de tecnologías bajas en carbono, se creó una reserva de 300 millones de permisos en el New Entrants Reserve (NER 300), un fondo destinado a financiar proyectos pioneros en energías renovables y tecnologías de captura y almacenamiento de carbono.

Durante esta fase, el mercado de carbono europeo experimentó un crecimiento significativo en volumen de transacciones, consolidando al EU ETS como uno de los principales impulsores del

mercado internacional de carbono. A pesar de los avances logrados, el exceso de permisos en circulación provocó una caída en el precio del carbono, esto subrayó la necesidad de crear mecanismos de ajuste de oferta, y así surgió la Reserva de Estabilidad del Mercado (MSR), que comenzaría a operar en la fase siguiente para gestionar el suministro de permisos y evitar grandes fluctuaciones de precio [34].

1.5. Marco Legislativo y Reformas del EU ETS

1.5.1. Fundamentos legales de la directiva ETS

El marco legal del EU ETS se basa en la Directiva 2003/87/EC, que instauró el primer sistema internacional de comercio de emisiones para combatir el cambio climático [35]. Este marco proporciona una estructura para gestionar y reducir las emisiones de gases de efecto invernadero (GEI) en sectores clave de la economía europea. La Directiva ETS fija las pautas para el funcionamiento de este sistema y establece los requisitos para que las instalaciones industriales y de generación de energía monitoricen y reporten sus emisiones, asegurando así la transparencia y el cumplimiento en todos los Estados Miembros. Desde su lanzamiento en 2005, la Directiva ha sido objeto de múltiples revisiones para fortalecer su efectividad y adaptarse a los compromisos climáticos de la Unión Europea, especialmente a la luz del Acuerdo de París y la Ley Climática Europea, que exige la neutralidad de carbono para 2050.

Cada fase del EU ETS ha traído consigo ajustes legislativos, diseñados para perfeccionar el sistema en aspectos como el método de asignación de permisos, la inclusión de nuevos sectores y el control de la oferta en el mercado de carbono. Estas revisiones han consolidado al EU ETS como una herramienta central de la política climática de la UE, al tiempo que han permitido el desarrollo de una infraestructura reguladora robusta para gestionar el comercio de emisiones [36].

1.5.2. Principales revisiones y ajustes a la directiva ETS

Desde sus primeras etapas, el EU ETS ha evolucionado mediante reformas que buscaban fortalecer su impacto. La introducción de la subasta de permisos en la tercera fase (2013-2020) promueve

una competencia justa entre los Estados Miembros y minimiza la distorsión del mercado. Otro cambio clave fue la implementación de la Reserva de Estabilidad del Mercado (MSR) en 2018, creada para ajustar automáticamente la oferta de permisos en el mercado y así mitigar los efectos de fluctuaciones económicas o crisis, como la de 2008, que redujo drásticamente las emisiones y acumuló un exceso de permisos en circulación [37]. Estos ajustes estructurales han sido esenciales para mantener la estabilidad del precio del carbono y asegurar que el EU ETS continúe proporcionando incentivos económicos para la reducción de emisiones.

1.5.3. Cambios “Fit for 55” de 2023

El paquete “Fit for 55”, adoptado en 2023, es una actualización significativa de la Directiva ETS que tiene como objetivo alinear el EU ETS con la meta de reducir al menos un 55% de las emisiones de GEI para 2030 respecto a los niveles de 1990. Esta revisión introduce ajustes profundos, incluyendo un aumento del factor de reducción anual de permisos, que pasa a 4.3% entre 2024 y 2027, y a 4.4% a partir de 2028, acelerando la disminución del límite total de emisiones [38]. Estos cambios buscan lograr una reducción del 62% para 2030 en comparación con 2005, consolidando el papel del EU ETS como un mecanismo central para alcanzar la neutralidad climática de la UE en 2050.

Entre las novedades más destacadas del paquete “Fit for 55” se encuentra la creación del ETS2, un segundo sistema de comercio de emisiones dirigido a sectores previamente excluidos del EU ETS, como el transporte por carretera, los edificios y otras fuentes de emisión más pequeñas. El ETS2, que entrará en funcionamiento en 2027, cubrirá sectores adicionales en los que hasta ahora la UE había aplicado enfoques regulatorios diferentes. Esto asegura una cobertura más amplia del mercado de carbono en toda la economía europea, permitiendo un mayor progreso en la reducción de emisiones en áreas donde las medidas anteriores habían sido limitadas.

Adicionalmente, se estableció el Fondo Social Climático (SCF), el cual movilizará hasta 86.7 mil millones de euros entre 2026 y 2032 para ayudar a los sectores sociales más vulnerables a enfrentar los efectos de los precios del carbono en el ETS2. Este fondo se destina a mitigar los impactos sociales derivados del coste del carbono en estos sectores, ofreciendo ayudas a hogares de bajos

ingresos y pequeñas empresas que puedan verse afectadas por el aumento de los precios de energía y transporte como consecuencia del ETS2.

Finalmente, el “Fit for 55” refuerza el papel de la Reserva de Estabilidad del Mercado (MSR), asegurando que esta continúe equilibrando la oferta de permisos a medida que el límite de emisiones se vuelve más estricto. Este paquete de reformas también modifica la asignación gratuita de permisos, haciéndola más restrictiva para sectores como la aviación, que dejará de recibir permisos gratuitos a partir de 2026, mientras que otros sectores verán reducidas sus asignaciones gratuitas en función de sus avances en la descarbonización. Estas reformas crean un marco de incentivos que refuerzan la competitividad de las empresas más sostenibles y alinean las políticas climáticas de la UE con sus compromisos internacionales.

1.6. Impacto y Objetivos Climáticos del EU ETS

1.6.1. Metas de reducción para 2030 y 2050

El EU ETS juega un papel crucial en la estrategia de la UE para alcanzar los objetivos de la Ley Climática Europea, que establecen la meta de lograr la neutralidad climática para 2050. Como primer hito importante, la UE se ha comprometido a reducir sus emisiones netas de gases de efecto invernadero en al menos un 55% para 2030 en comparación con los niveles de 1990 [39]. Esta ambiciosa meta de reducción, en línea con el Acuerdo de París, exige una transformación estructural en la manera en que los sectores cubiertos por el EU ETS gestionan sus emisiones.

La contribución del EU ETS es significativa porque establece un límite de emisiones decreciente cada año, ajustado según un factor de reducción que aumentó a un 4.3% en el periodo 2024-2027 y luego a un 4.4% anual a partir de 2028. Estas disminuciones continuas y progresivas del límite de emisiones facilitan que la UE mantenga una trayectoria de reducción sostenida, incentivando a las industrias a adoptar tecnologías y prácticas más limpias para adaptarse a las nuevas realidades del mercado de carbono [40]. A través de este límite decreciente, el EU ETS ha logrado reducir las emisiones de los sectores cubiertos en un 47% desde su creación en 2005, consolidándose como un mecanismo central en la política climática de la UE y estableciendo un precedente para otros sistemas de comercio de emisiones.

1.6.2. Inversión en innovación y modernización

Los ingresos generados por la subasta de permisos en el EU ETS se destinan a iniciativas orientadas a fomentar tecnologías bajas en carbono y mejorar la eficiencia energética, contribuyendo al objetivo de la UE de descarbonizar su economía. Desde 2013, el EU ETS ha generado más de 200 mil millones de euros en ingresos, los cuales han sido destinados en gran medida a apoyar inversiones en innovación y modernización. Los Estados Miembros están obligados a invertir al menos el 50% de estos ingresos en proyectos relacionados con la acción climática y la transición energética, como el desarrollo de energías renovables, la movilidad sostenible y la eficiencia energética en sectores clave [41].

Dos fondos esenciales financiados por los ingresos del EU ETS son el Fondo de Innovación y el Fondo de Modernización. El Fondo de Innovación se enfoca en promover tecnologías pioneras en áreas de energía limpia, como la captura y almacenamiento de carbono, la producción de hidrógeno verde y soluciones avanzadas para reducir las emisiones industriales. Este fondo permite que los proyectos de alto impacto en descarbonización tengan acceso a recursos significativos, fomentando la adopción de tecnologías que reduzcan la dependencia de los combustibles fósiles y permitan una transición hacia una industria más limpia.

Por su parte, el Fondo de Modernización tiene como objetivo facilitar el acceso a financiamiento para que los Estados Miembros con economías en transición modernicen sus sistemas energéticos. Este fondo ofrece recursos para mejorar la infraestructura energética en sectores como la generación de electricidad y calefacción, promoviendo la eficiencia y la reducción de emisiones. El Fondo de Modernización se ha dirigido principalmente a Estados Miembros de Europa Central y Oriental, donde el avance hacia una economía baja en carbono es crucial para reducir la dependencia de fuentes de energía intensivas en carbono y para asegurar una transición justa en toda la región.

Además, con la expansión del EU ETS en sectores adicionales, como edificios y transporte por carretera a través del nuevo ETS2, los ingresos se utilizarán también para financiar el Fondo Social Climático (SCF). Este fondo aborda los posibles impactos sociales del aumento en el precio del

carbono sobre los ciudadanos más vulnerables, asegurando que la transición hacia una economía baja en carbono sea inclusiva y equitativa.

1.7. Integridad y Supervisión del Mercado de Carbono

1.7.1. Supervisión y Prevención de Abusos

La integridad del mercado de carbono es crucial para garantizar que el EU ETS funcione de manera justa y transparente, evitando conductas especulativas o manipulativas que puedan distorsionar los precios del carbono. La supervisión del mercado está diseñada para mantener una operativa alineada con los principios de seguridad y confianza. La Comisión Europea, en colaboración con la Autoridad Europea de Valores y Mercados (ESMA), examina regularmente los comportamientos de trading para identificar cualquier riesgo potencial de manipulación o prácticas injustas [6]. Esto incluye la implementación de estándares que limitan prácticas como la difusión de información falsa o el uso de información privilegiada, que podrían generar un aumento artificial del precio de los permisos o afectar la estabilidad del mercado. Los resultados de los análisis de ESMA han confirmado que el mercado de carbono de la UE se mantiene ordenado y comparable con otros mercados financieros, sin evidencia de abuso de mercado hasta la fecha.

Para detectar y abordar cualquier irregularidad, el sistema de transacciones del EU ETS también se beneficia del registro y control de todas las transacciones a través del European Union Transaction Log (EUTL). Este sistema de monitoreo centralizado revisa automáticamente cada transacción para garantizar que se realice de acuerdo con las normativas establecidas, fortaleciendo así la transparencia y la confianza en el mercado [42].

1.7.2. Regulaciones financieras

Desde 2018, el EU ETS se encuentra bajo la supervisión de un marco regulador financiero robusto, lo cual le permite funcionar con altos estándares de integridad comparables a los aplicados en otros mercados financieros de la UE. Las Directivas sobre Mercados de Instrumentos Financieros II (MiFID2) y la Regulación sobre Abuso de Mercado (MAR) son componentes esenciales de este

marco regulador. La clasificación de los permisos de emisión como instrumentos financieros según MiFID2 implica que el EU ETS debe cumplir con un conjunto estricto de normas de integridad, transparencia y lucha contra el abuso de mercado. Esto significa que todos los participantes en el mercado de carbono deben someterse a controles y salvaguardias contra actividades fraudulentas, como el lavado de dinero y el uso de información privilegiada [43].

La MAR, por su parte, prohíbe explícitamente prácticas como la manipulación de mercado, estableciendo medidas que limitan la posibilidad de que se generen fluctuaciones anormales en los precios de los permisos. Asimismo, se exige a las empresas de alto nivel de emisión (con instalaciones importantes en el sistema) que gestionen la información interna de manera rigurosa, evitando la filtración de datos sensibles que puedan beneficiar a participantes específicos. A esto se suman normativas de transparencia que facilitan el acceso a información pública sobre las operaciones de comercio de permisos, lo cual permite que todos los actores, independientemente de su tamaño o acceso al mercado, tengan visibilidad sobre precios y volúmenes negociados.

Adicionalmente, el marco regulador abarca otras disposiciones, como la Directiva contra el Lavado de Dinero (AMLD), que impone requisitos para identificar a los participantes y verificar la legitimidad de sus fondos. En conjunto, estas regulaciones financieras no solo previenen el abuso de mercado y aseguran una competencia equitativa, sino que también consolidan la posición del EU ETS como un mercado transparente y seguro, alineado con los objetivos climáticos de la UE y los estándares de buen gobierno en los mercados de carbono.

1.8. Mercado Internacional de Carbono y Cooperación

1.8.1. Colaboración global

El EU ETS ha promovido la colaboración internacional en el desarrollo de mercados de carbono, reconociendo que un esfuerzo global puede maximizar el impacto en la reducción de emisiones. En particular, la UE ha trabajado de cerca con China y Corea en proyectos de cooperación técnica y asistencia para el diseño de sus sistemas de comercio de emisiones, como el “Korean Emissions Trading System” (KETS) y el sistema nacional chino. Estos esfuerzos han permitido la transferencia de conocimiento en metodologías de monitoreo, reporte y verificación, así como en

la estructura de los mercados de permisos de emisión [44]. La colaboración con China, a través de un proyecto que se extendió hasta 2024, fomentó la creación de un sistema de comercio a nivel nacional, además de ofrecer capacitación para funcionarios y establecer canales de diálogo técnico entre ambas regiones. Estos intercambios han sido fundamentales para que sistemas como el de China puedan avanzar hacia una implementación alineada con las mejores prácticas globales, sentando bases para una mayor armonización [45].

1.8.2. Acuerdos bilaterales y multilaterales

El Acuerdo de París proporciona un marco robusto para el uso de mercados internacionales de carbono y establece mecanismos como el artículo 6, que permite a las partes intercambiar unidades de carbono para cumplir con sus objetivos de reducción de emisiones. Este mecanismo está diseñado para asegurar una contabilidad común y transparente en los mercados, lo cual es esencial para evitar la doble contabilidad y asegurar que los objetivos climáticos sean realmente alcanzados. En este contexto, la UE ha firmado acuerdos bilaterales para enlazar su sistema de comercio de emisiones con otros sistemas nacionales, un ejemplo destacado es el acuerdo con Suiza en 2020, que permite la aceptación mutua de permisos en ambos sistemas. Este acuerdo fue posible gracias a la compatibilidad en los estándares de monitoreo y reporte y establece un precedente para futuros enlaces con sistemas que compartan un compromiso de reducción absoluta de emisiones y un sistema de comercio compatible.

Además de los acuerdos bilaterales, la UE es miembro fundador de la International Carbon Action Partnership (ICAP), una plataforma que reúne a jurisdicciones con sistemas de límite y comercio de emisiones. Esta asociación facilita el intercambio de experiencias y promueve el desarrollo de capacidades mediante entrenamientos y talleres regulares, permitiendo que los miembros aprendan de los desafíos y logros de otros sistemas. La UE también participa en la “Partnership for Market Readiness” del Banco Mundial, que apoya a países en desarrollo en la preparación de instrumentos de mercado para reducir sus emisiones, y en la “Carbon Market Platform” de la OCDE, que busca expandir el alcance de los mercados de carbono y fomentar un terreno de juego equitativo en la fijación de precios de carbono a nivel global.

1.8.3. Distribución global de instrumentos de precios al carbono

Aunque la Unión Europea fue pionera en establecer un sistema de comercio de emisiones con el EU ETS en 2005, en los últimos años otros países y regiones han desarrollado sus propios instrumentos de precios de carbono, expandiendo significativamente el alcance global de estas políticas. Este mapa ilustra cómo distintas jurisdicciones han implementado mecanismos de fijación de precios al carbono, incluyendo sistemas de comercio de emisiones y créditos de carbono, que abarcan una parte importante de las emisiones globales.

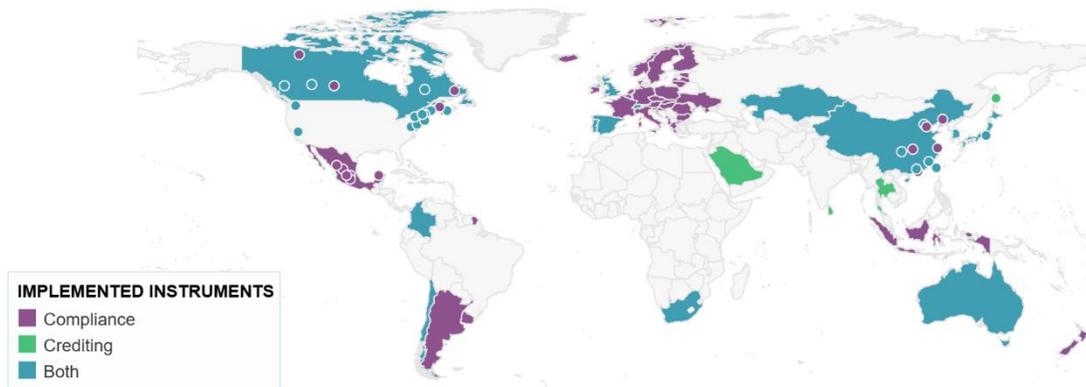


Figura 2: Instrumentos de Fijación de Precios de Carbono Implementados a Nivel Global en 2024 [46]

La adopción de estos sistemas alrededor del mundo, como los implementados en Canadá, China, y varias naciones latinoamericanas, demuestra el compromiso creciente para reducir emisiones de manera costo-efectiva y apoyar la transición hacia una economía baja en carbono. Esta expansión refleja el reconocimiento de que los mercados de carbono pueden jugar un rol crucial en la mitigación del cambio climático, promoviendo la colaboración y estableciendo un estándar para futuras iniciativas.

1.9. Diplomacia en Precios Internacionales de Carbono

1.9.1. Taskforce para el precio internacional de carbono

La Taskforce para la Diplomacia de Precios Internacionales de Carbono, lanzada por la Comisión Europea en 2024, se dedica a promover y expandir el uso de mercados de carbono a nivel mundial.

Este esfuerzo tiene como objetivo apoyar a los países en la adopción de políticas de precios de carbono alineadas con los objetivos del Acuerdo de París [47]. La iniciativa trabaja en tres pilares fundamentales: la expansión de mercados de carbono nacionales y regionales, el desarrollo de mercados de cumplimiento bajo el artículo 6 del Acuerdo de París y la promoción de altos estándares de integridad en los mercados voluntarios de carbono.

Uno de los objetivos centrales de la Taskforce es aumentar la cobertura global de emisiones con precios de carbono para contribuir de manera significativa a la reducción de gases de efecto invernadero. La UE fomenta la adopción de estos precios de carbono entre sus socios internacionales, destacando cómo la fijación de precios puede reducir las emisiones de forma costo-efectiva y generar ingresos para financiar la transición verde en cada país [48]. La iniciativa apoya también el “Global Carbon Pricing Challenge” de Canadá, que aboga por que el 60% de las emisiones globales estén cubiertas por un precio de carbono para 2030.

La Taskforce también desempeña un papel en la supervisión y apoyo de la implementación de los requisitos del artículo 6 del Acuerdo de París, promoviendo la transparencia en las transacciones de carbono internacionales. Esto incluye la asistencia en la creación de sistemas de contabilidad robustos para los mercados de cumplimiento internacionales y la colaboración en la definición de estándares globales para los mercados voluntarios de carbono, fortaleciendo así la confianza en estos mecanismos y asegurando que contribuyan efectivamente a los objetivos climáticos globales.

2. Objetivos del trabajo

El presente trabajo tiene como objetivo principal analizar en profundidad las dinámicas del mercado de emisiones de carbono de la Unión Europea (EU ETS) y desarrollar modelos predictivos que permitan anticipar los precios de las emisiones de CO₂. Este análisis se fundamenta en la combinación de métodos estadísticos y algoritmos avanzados de machine learning, ofreciendo una visión integral del comportamiento de este mercado en el contexto de los esfuerzos por mitigar el cambio climático.

El estudio busca proporcionar una comprensión clara y precisa de los factores clave que influyen en el precio del carbono y evaluar la capacidad predictiva de diferentes enfoques. En este

sentido, se busca no solo explicar el comportamiento pasado del mercado, sino también avanzar en la implementación de herramientas que puedan contribuir al análisis de escenarios futuros.

A continuación, se detallan los objetivos que orientan el desarrollo de este trabajo:

2.1. Objetivo General

El objetivo principal del trabajo es analizar el mercado de emisiones de carbono del EU ETS desde una perspectiva cuantitativa, identificando los factores clave que afectan los precios de las emisiones y aplicando diversos modelos predictivos para evaluar su precisión y aplicabilidad en la predicción de precios futuros.

2.2. Objetivos Específicos

1. Estudio del funcionamiento y evolución del EU ETS:

Comprender el diseño, la evolución histórica y el impacto del EU ETS como el sistema más avanzado de comercio de emisiones a nivel global. Esto incluye el análisis de sus componentes fundamentales, como las fases de implementación, las reformas legislativas y las dinámicas de mercado.

2. Identificación de las variables clave:

Analizar un conjunto de variables que influyen en el mercado de carbono, incluyendo factores macroeconómicos, indicadores energéticos y variables específicas del EU ETS, con el fin de determinar aquellas que tienen mayor impacto en la formación de precios.

3. Implementación de modelos predictivos:

Aplicar y evaluar diferentes enfoques predictivos, desde métodos estadísticos tradicionales hasta técnicas avanzadas de machine learning, para analizar su capacidad de modelar la compleja interacción entre las variables del mercado.

4. Validación y comparación de modelos:

Comparar el rendimiento de los modelos predictivos implementados utilizando métricas estándar como RMSE, MSE y R^2 . Este análisis busca identificar el modelo que mejor se adapta a los datos disponibles y que ofrece las predicciones más precisas.

5. Análisis de resultados y reflexiones finales:

Interpretar los resultados obtenidos de los modelos predictivos en función de su aplicabilidad y utilidad para el análisis del mercado de carbono, destacando sus implicaciones en el contexto de las políticas climáticas de la Unión Europea.

2.3. Relevancia de los Objetivos

El mercado de emisiones de carbono juega un papel central en la estrategia de descarbonización de la Unión Europea, y su correcto funcionamiento es crucial para alcanzar los objetivos climáticos establecidos en el Acuerdo de París y el Pacto Verde Europeo. Este trabajo tiene como propósito aportar un análisis riguroso y técnicas avanzadas de modelado para ofrecer una comprensión más profunda de este sistema, contribuyendo al conocimiento académico y práctico sobre su comportamiento.

A través de la identificación de las variables más influyentes y la evaluación de modelos predictivos, el trabajo busca llenar un vacío en la investigación actual, donde la predicción precisa de los precios del carbono sigue siendo un desafío técnico y metodológico. Además, al centrarse en un enfoque exclusivamente analítico y predictivo, este estudio se alinea directamente con las necesidades actuales de formulación de políticas basadas en evidencia.

2.4. Preguntas Clave del Trabajo

Para cumplir con los objetivos planteados, el trabajo se estructura en torno a las siguientes preguntas clave:

- ¿Cuáles son las variables más relevantes para explicar el precio del carbono en el EU ETS?

- ¿Qué técnicas predictivas son más efectivas para anticipar el comportamiento del mercado de emisiones?
- ¿Qué patrones se pueden identificar en el precio del carbono utilizando técnicas de machine learning?
- ¿Cómo se pueden interpretar los resultados del modelado en el contexto de las políticas climáticas de la UE?

3. Herramientas y Técnicas de Programación para el Análisis y Modelado de Datos

3.1. Python

Python es un lenguaje de programación interpretado, de alto nivel y de propósito general, creado por Guido van Rossum y lanzado en 1991. Diseñado con una sintaxis limpia y legible, Python facilita la escritura de código claro y organizado. Su popularidad se debe, en gran medida, a su enfoque en la simplicidad, haciéndolo accesible tanto para principiantes como para desarrolladores experimentados.

3.1.1. Historia de Python

Python fue creado en los Países Bajos por van Rossum, quien buscaba un lenguaje fácil de entender y con un enfoque en la productividad del programador. El nombre “Python” proviene del grupo de comedia británico Monty Python, reflejando la intención de crear un lenguaje “divertido” de usar. A lo largo de los años, Python ha evolucionado a través de varias versiones. La introducción de Python 2 en 2000 trajo características como “list comprehensions” y una recolección de basura más robusta. En 2008, se lanzó Python 3, una versión importante que mejoró la consistencia y eficiencia del lenguaje, aunque no fue retrocompatible, lo cual requirió tiempo para que la comunidad migrara sus proyectos.

3.1.2. Filosofía y ventajas de Python

Python se basa en el principio de “legibilidad del código”, como se ve en el “Zen de Python”, una serie de aforismos que reflejan la filosofía del lenguaje, entre los cuales se encuentran: “Lo explícito es mejor que lo implícito” y “La simplicidad es mejor que la complejidad”. La sintaxis de Python es concisa, evitando corchetes o puntos y comas innecesarios, lo cual permite que los programadores enfoquen más su atención en la lógica del código que en su estructura.

3.1.3. Ecosistema y bibliotecas de Python

El ecosistema de Python es uno de sus mayores puntos fuertes. Existen bibliotecas especializadas para casi cualquier aplicación, como NumPy y Pandas para análisis de datos, Matplotlib y Seaborn para visualización, Scikit-Learn para machine learning, y TensorFlow y PyTorch para aprendizaje profundo. Este amplio ecosistema hace que Python sea una elección natural para proyectos de ciencia de datos, inteligencia artificial y modelado predictivo.

3.1.4. Comunidad y soporte

La comunidad de Python es enorme y global, con usuarios en prácticamente cualquier campo de la tecnología y la ciencia. Existen foros, documentación extensa y numerosos cursos gratuitos y de pago que facilitan su aprendizaje. La comunidad abierta también contribuye al desarrollo y mejora continua del lenguaje, asegurando que esté actualizado con las últimas tendencias y necesidades tecnológicas.

3.1.5. Aplicaciones en ciencia de datos y aprendizaje automático

En el ámbito de la ciencia de datos, Python se ha convertido en el lenguaje preferido debido a su simplicidad y potencia. Su uso en análisis de datos y machine learning permite a los científicos de datos realizar desde tareas básicas de procesamiento y limpieza hasta la implementación de complejos modelos predictivos. La integración con bibliotecas avanzadas permite a los usuarios trabajar con grandes volúmenes de datos, desarrollar modelos de clasificación y regresión, y

realizar análisis de series temporales, como los que se necesitan en el análisis del mercado de derechos de emisión.

Python también destaca por su capacidad de trabajar en entornos como Jupyter Notebooks, donde los analistas pueden escribir código, visualizar gráficos y documentar su proceso en un solo lugar, lo cual es ideal para investigación y colaboración en proyectos de ciencia de datos.

3.2. Jupyter Notebooks

Jupyter Notebooks es una herramienta de código abierto ampliamente utilizada en ciencia de datos y análisis de datos. Creado por el proyecto Jupyter, este entorno permite combinar código, texto, gráficos y visualizaciones en un solo documento interactivo, facilitando el desarrollo, la documentación y la presentación de análisis complejos. Originalmente diseñado para lenguajes como Python, Julia y R, hoy en día Jupyter se ha convertido en una plataforma esencial para programadores y científicos de datos, pues ofrece flexibilidad y accesibilidad en el flujo de trabajo [49].

3.1.1. Características clave de Jupyter Notebooks

Uno de los aspectos más distintivos de Jupyter Notebooks es su capacidad de ejecutar código en celdas individuales, lo que permite a los usuarios probar, modificar y documentar sus experimentos paso a paso. Esta estructura facilita la edición y experimentación sin necesidad de ejecutar todo el código en cada cambio, lo cual es particularmente ventajoso para proyectos de aprendizaje automático y análisis de grandes volúmenes de datos. Además, Jupyter Notebooks soporta Markdown, permitiendo enriquecer los documentos con secciones de texto, fórmulas y explicaciones detalladas para facilitar la interpretación y presentación de los resultados.

3.1.2. Ventajas para la ciencia de datos y el aprendizaje automático

La posibilidad de visualizar gráficos y tablas directamente junto al código y texto explicativo convierte a Jupyter Notebooks en una herramienta ideal para el análisis exploratorio de datos (EDA)

y para la construcción de modelos predictivos, como los que se implementarán en este proyecto. Los usuarios pueden integrar bibliotecas de visualización, como Matplotlib o Seaborn, y realizar visualizaciones interactivas con herramientas como Plotly. Esta capacidad de mostrar gráficos y resultados en tiempo real facilita la comprensión de patrones y tendencias en los datos, elementos esenciales en la predicción de precios de mercado y análisis de series temporales.

3.1.3. Jupyter como plataforma de colaboración

Jupyter Notebooks también es una herramienta eficaz para la colaboración en ciencia de datos. Al combinar el análisis técnico con explicaciones documentadas, Jupyter permite a los equipos de trabajo, incluso a aquellos con niveles de experiencia diferentes, compartir y revisar proyectos de forma intuitiva. Los notebooks pueden guardarse y compartirse en plataformas como GitHub, facilitando la colaboración en proyectos a nivel académico e industrial. Además, la opción de exportar los notebooks a otros formatos (PDF, HTML) contribuye a la flexibilidad en la presentación de resultados.

3.1.4. Aplicación en el análisis de datos del EU ETS

Para el presente análisis del mercado de emisiones de carbono del EU ETS, Jupyter Notebooks proporcionará un entorno estructurado y eficaz para desarrollar los modelos predictivos, facilitando la visualización de series temporales, la interacción con datos históricos y la documentación detallada de cada paso en el proceso de modelado. Esta herramienta será clave para combinar análisis de series temporales, algoritmos de aprendizaje automático y métodos de regresión en un formato claro y ordenado.

Nota sobre el Uso de Inteligencia Artificial Generativa

En el desarrollo del presente trabajo, se emplearon herramientas de Inteligencia Artificial Generativa (IAG) como apoyo en tareas de optimización de redacción y traducción. Todas las decisiones finales respecto al contenido, análisis y conclusiones del presente trabajo han sido

realizadas por el autor, garantizando su alineación con los objetivos del proyecto y los principios académicos de rigor y originalidad.

4. Estudio de las Variables

Este capítulo se centra en la base conceptual y analítica del estudio, comenzando con una descripción detallada de las variables recopiladas para el modelo predictivo. Estas variables, que abarcan aspectos regulatorios, económicos, industriales y energéticos, son esenciales para comprender las dinámicas del Sistema de Comercio de Emisiones de la Unión Europea (EU ETS) y su influencia en el precio de las emisiones de CO₂. Además de su relevancia individual, estas variables reflejan tendencias más amplias en la economía europea, el mercado energético y las políticas climáticas.

La primera parte del capítulo organiza las variables en grupos según su naturaleza y función, proporcionando una estructura clara que facilita tanto el análisis como la interpretación de los resultados. Este enfoque permite identificar cómo cada conjunto de variables contribuye al objetivo principal del modelo y destaca aquellas que presentan una relación directa con el precio de las emisiones o representan factores clave en el contexto regulatorio.

Posteriormente, se llevará a cabo un análisis de correlaciones para evaluar las relaciones entre las variables recopiladas. Este análisis es crucial para identificar posibles redundancias y patrones de colinealidad que puedan afectar la estabilidad y precisión del modelo. Al priorizar las variables más relevantes, se logra no solo una mayor eficiencia computacional, sino también una mejora en la capacidad interpretativa y predictiva del modelo.

La última sección del capítulo se dedicará al análisis exploratorio de los datos, utilizando visualizaciones y estadísticas descriptivas para identificar patrones, tendencias y anomalías presentes en el conjunto de datos. Este análisis permitirá obtener una comprensión más profunda de las dinámicas subyacentes y establecer relaciones preliminares entre las variables independientes y la variable objetivo. Especial atención se dedicará a las variables binarias que reflejan decisiones regulatorias, para evaluar su impacto específico en el precio de las emisiones.

En su conjunto, este capítulo proporciona una base sólida para los análisis avanzados que se desarrollarán en las secciones siguientes, estableciendo una comprensión integral de las variables y su comportamiento en el contexto del modelo de predicción.

4.1. Presentación y Clasificación de las Variables

En esta sección, presentamos la lista completa de variables utilizadas en el análisis, organizadas en grupos según su naturaleza o función. Este enfoque estructurado facilita la comprensión de cómo cada grupo de variables contribuye al objetivo principal: modelar y predecir el precio de la tonelada de CO₂ dentro del Sistema de Comercio de Emisiones de la Unión Europea (EU ETS). A continuación, se detallan las categorías y el propósito de cada grupo.

1. Fechas y Precios:

Este grupo incluye variables esenciales para organizar los datos temporalmente y estructurar el análisis:

- **Date:** Marca la fecha asociada a cada registro, proporcionando una referencia temporal fundamental para el análisis de tendencias y estacionalidad.
- **Auction Price €/tCO₂:** Es la variable objetivo del análisis. Representa el precio de subasta de una tonelada de CO₂ dentro del EU ETS.

Estas variables no actúan como predictoras, pero son fundamentales para el marco temporal y la definición del objetivo.

2. Estadísticas del Precio:

Estas variables derivan directamente del precio de subasta y caracterizan su distribución en el mercado:

- **Minimum Bid €/tCO₂:** Precio mínimo registrado en las ofertas de la subasta.
- **Maximum Bid €/tCO₂:** Precio máximo registrado en las ofertas de la subasta.
- **Mean €/tCO₂:** Promedio de los precios registrados en la subasta.
- **Median €/tCO₂:** Mediana de los precios registrados en la subasta.

Aunque estas variables son descriptivas y no se incluyen como predictores independientes, ayudan a comprender la dinámica del precio.

3. Variables de Subastas:

Este grupo refleja las dinámicas específicas del mecanismo de subastas dentro del EU ETS:

- **Auction Volume tCO₂:** Cantidad total de permisos de emisión disponibles en cada subasta.
- **Total Amount of Bids:** Número total de ofertas presentadas en cada subasta.
- **Cover Ratio:** Relación entre la cantidad de permisos demandados y la cantidad ofertada.
- **Total Revenue €:** Ingresos totales generados por la subasta.

Estas variables capturan la interacción entre oferta y demanda, proporcionando contexto sobre las condiciones del mercado.

5. Regulaciones Ambientales:

Incluye variables binarias que reflejan la implementación de políticas climáticas clave en la Unión Europea:

- **European Green Deal**
- **European Climate Law**
- **Fit for 55 Package**
- **Market Stability Reserve (MSR)**
- **Carbon Border Adjustment Mechanism (CBAM)**
- **2040 Climate Target**
- **EU Adaptation Strategy**

Estas variables representan cambios estructurales en el mercado derivados de nuevas regulaciones, que influyen en la percepción del precio y las expectativas del mercado.

5. Indicadores Macroeconómicos:

Incluye factores económicos generales que influyen indirectamente en el precio del carbono:

- **Emissions:** Total de emisiones reportadas, una métrica relacionada directamente con la necesidad de permisos.
- **GDP:** Producto Interno Bruto, un indicador de la actividad económica general.

- **Interest Rates:** Tasas de interés, relevantes para entender las condiciones monetarias.
- **Inflation Rate:** Refleja la estabilidad económica y el poder adquisitivo.
- **MSCI Europe:** Índice que mide el rendimiento de las principales empresas europeas.
- **USD/EUR Exchange Rate:** Tasa de cambio entre el dólar y el euro, clave para transacciones internacionales.

Estas variables permiten contextualizar el mercado de carbono dentro del entorno económico más amplio.

6. Variables Energéticas:

Agrupar variables clave relacionadas con el sector energético, responsable de una proporción significativa de emisiones:

- **Electricity Consumption by Country:** Contiene el consumo eléctrico de 24 países europeos. Esta categoría general engloba las siguientes variables:
 - *Austria, Belgium, Bulgaria, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland.*
- **Gas Price (USD/MMBtu):** Precio del gas natural, un indicador clave de los costos energéticos.
- **Oil Price:** Precio del petróleo, otro factor crucial para los costos de energía.
- **Renewables (% equivalent primary energy):** Proporción de energías renovables utilizadas como equivalente de energía primaria.

Este grupo es fundamental para comprender las presiones del mercado energético sobre el precio de las emisiones.

7. Indicadores Industriales:

Incluye variables específicas de sectores económicos intensivos en emisiones:

- **Producción Industrial:** Refleja el nivel general de actividad económica en los sectores industriales.

- **Steel Production:** Indicador específico de un sector clave y altamente emisor como la producción de acero.

Estas variables permiten conectar la actividad económica e industrial con la dinámica del mercado de emisiones.

4.1.1. Fuentes de datos y métodos de recopilación

La sección de fuentes de datos detalla las bases de datos y recursos utilizados para recopilar la información necesaria para el análisis y modelado. La selección de estas fuentes garantiza la fiabilidad y actualidad de las variables analizadas en el marco del Sistema de Comercio de Emisiones de la Unión Europea (EU ETS). A continuación, se describen las fuentes específicas utilizadas para cada categoría de variables, en el mismo orden en que fueron presentadas:

1. Fechas y Precios:

Las fechas y precios de subasta provienen directamente de los datos oficiales publicados por la European Energy Exchange (EEX), el mercado autorizado para gestionar las subastas del EU ETS. Este recurso proporciona:

- Fechas específicas de las subastas.
- Precios de subasta por tonelada de CO₂, incluyendo precios mínimos, máximos y el promedio registrado en cada evento.
- **Fuente:** European Energy Exchange.

2. Estadísticas del Precio:

Los datos estadísticos relacionados con los precios de subasta también se obtuvieron de la EEX, que proporciona una visión detallada de las dinámicas de precios, incluyendo medianas y otros indicadores clave. Estas estadísticas complementan el análisis al caracterizar la distribución del precio de las subastas. **Fuente:** European Energy Exchange.

3. Variables de Subastas

La información sobre volumen de subastas, número de ofertas, "cover ratio" e ingresos totales generados se recopiló igualmente de la plataforma de la EEX. Estos datos reflejan la interacción

entre oferta y demanda en las subastas de permisos de emisión. **Fuente:** European Energy Exchange.

4. Regulaciones Ambientales

Las políticas ambientales y los cambios regulatorios relevantes, como el European Green Deal, la Market Stability Reserve (MSR) o el Carbon Border Adjustment Mechanism (CBAM), se documentaron a través de informes y comunicados de la Comisión Europea y otras fuentes gubernamentales. Estas políticas afectan la percepción del mercado y están disponibles en recursos oficiales:

Fuente: Comisión Europea.

5. Indicadores Macroeconómicos

Los datos macroeconómicos, como el PIB, tasas de interés, inflación y tipo de cambio USD/EUR, se obtuvieron de las siguientes fuentes:

- PIB e indicadores económicos generales: Eurostat.
- Tasas de interés: Banco Central Europeo.
- Tipo de cambio USD/EUR: FRED Economic Data.
- Índice MSCI Europe: Curvo MSCI Europe.

6. Variables Energéticas

El consumo eléctrico por país se obtuvo a través de bases de datos de la Agencia Europea de Medio Ambiente (EEA), mientras que los precios de gas y petróleo se recopilaron de plataformas de análisis energético global. Además, la proporción de energías renovables en el consumo energético se tomó de fuentes de análisis estadístico europeo:

- Consumo eléctrico: European Environment Agency.
- Precios de gas y petróleo: MacroTrends.
- Energías renovables: Eurostat.

7. Indicadores Industriales

Los datos relacionados con la producción industrial y de acero se recopilaron de fuentes sectoriales y estadísticas oficiales:

- Producción de acero: European Steel Association (EUROFER).
- Producción industrial: Eurostat.

4.2. Análisis de Correlaciones y Selección de Variables

Una vez identificadas todas las variables recopiladas para construir el modelo de predicción, el siguiente paso fundamental consiste en analizar las correlaciones existentes entre ellas. Este análisis es crucial para seleccionar las variables más relevantes y garantizar la robustez y precisión del modelo, evitando así problemas derivados de la redundancia de información. Cuando dos o más variables están altamente correlacionadas, comparten gran parte de la misma información, lo que genera un fenómeno conocido como colinealidad o multicolinealidad, dependiendo de si afecta a dos o múltiples variables, respectivamente. Este problema puede afectar negativamente al modelo de diversas maneras.

La presencia de colinealidad implica que las variables aportan información redundante, lo que no añade valor adicional al modelo. Esto no solo incrementa la complejidad innecesariamente, sino que también dificulta la interpretación de los resultados, ya que los efectos individuales de cada variable se confunden entre sí. En los modelos lineales o de regresión, este fenómeno puede causar inestabilidad en los coeficientes estimados, haciendo que pequeños cambios en los datos de entrenamiento resulten en variaciones significativas en los valores de los coeficientes. Esto compromete la confianza en las conclusiones obtenidas a partir del modelo.

Otro efecto negativo de la colinealidad es el incremento del riesgo de sobreajuste. Cuando el modelo incorpora demasiadas variables correlacionadas, tiende a ajustarse excesivamente a las particularidades del conjunto de datos de entrenamiento, perdiendo capacidad de generalización para datos nuevos. Además, la inclusión de variables redundantes también conlleva un mayor consumo de recursos computacionales, aumentando el tiempo y los costos asociados al desarrollo del modelo.

Para evitar estos problemas y optimizar el rendimiento del modelo, es necesario reducir el número de variables a aquellas que sean más relevantes para la tarea de predicción. Este proceso, conocido como selección de características, tiene como objetivo mejorar la interpretación del modelo al simplificarlo, evitar el sobreajuste y garantizar su eficiencia computacional. Una selección cuidadosa de las variables permite centrar el análisis en las que realmente tienen un impacto significativo en la variable objetivo, lo que mejora tanto la comprensión del modelo como su desempeño predictivo.

Para abordar la colinealidad y seleccionar adecuadamente las variables, se puede recurrir a diversas estrategias. Entre las técnicas más comunes se encuentran el análisis de correlación, que permite identificar pares de variables con alta correlación, y la eliminación de aquellas menos significativas dentro de cada grupo correlacionado. También es posible emplear transformaciones avanzadas como el análisis de componentes principales, que convierte un conjunto de variables correlacionadas en un número menor de componentes independientes que capturan la mayor parte de la variación en los datos originales.

Para realizar un análisis más efectivo de las variables recopiladas, se comenzó por excluir aquellas que están directamente relacionadas con el precio de la tonelada de CO₂, ya que estas no aportan valor como predictores independientes. Entre las variables eliminadas se encuentran indicadores como el precio mínimo, máximo, promedio y mediano de las subastas, así como variables relacionadas con el volumen de permisos, la cantidad de ofertas, la relación oferta-demanda y los ingresos totales generados por las subastas. Estas exclusiones garantizan que el análisis se centre únicamente en aquellas variables con potencial para contribuir a la predicción del precio de las emisiones de CO₂.

Antes de realizar un análisis detallado de correlaciones, queríamos evaluar cómo de correlacionados están los precios de electricidad de los diferentes países de la Unión Europea. El objetivo era determinar si existía una alta correlación entre ellos y, en ese caso, unificarlos en una única variable que represente el promedio de todas.

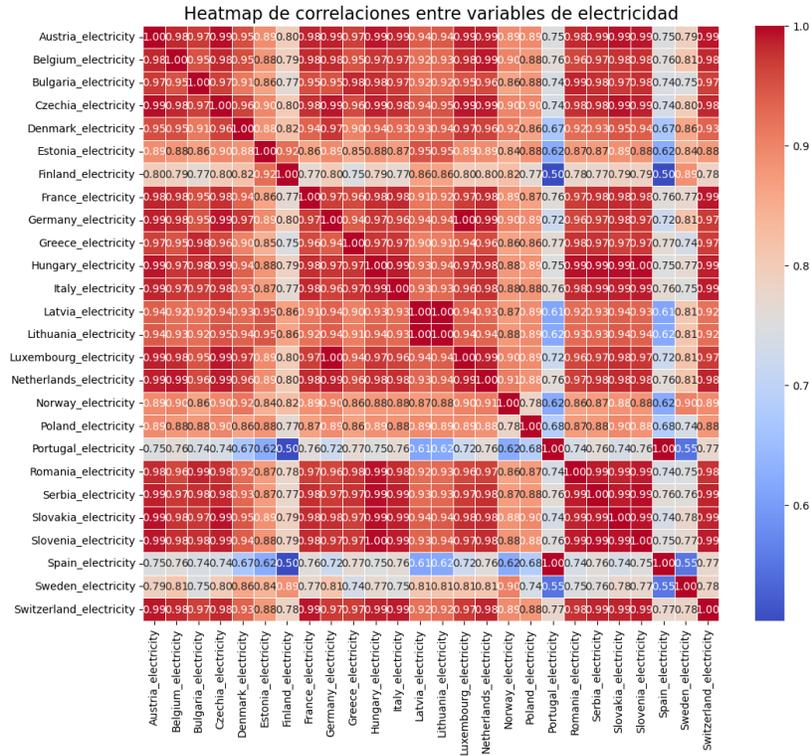


Figura 3: Mapa de calor de las variables precios de electricidad

Como se puede ver en el mapa de calor anterior, las correlaciones obtenidas mostraron que la mayoría de las variables relacionadas con los precios de electricidad de los países están altamente correlacionadas entre sí. Sin embargo, identificamos que las variables correspondientes a Portugal y España presentan una menor correlación con el resto de los países. Esto se debe, principalmente, a la alta proporción de energías renovables en la generación eléctrica de la península ibérica, así como a la gran cantidad de horas de luz solar que ambos países disfrutaron. Estas características hacen que sus precios de electricidad estén menos influenciados por factores comunes a otros países europeos, como los combustibles fósiles o la interdependencia de mercados eléctricos más centralizados. A pesar de estas diferencias, hemos decidido incluir estas variables en el análisis agregado, ya que aportan información valiosa sobre patrones particulares y excepciones dentro del conjunto de datos.

4.2.1. Reducción de dimensionalidad: Análisis de Componentes Principales (PCA)

Tras este filtrado inicial, se procedió a realizar un análisis de componentes principales (PCA, por sus siglas en inglés) para reducir la dimensionalidad del conjunto de datos y captar la mayor cantidad de información posible en un menor número de variables. Previo a la aplicación del PCA, los datos fueron estandarizados utilizando un escalador estándar (StandardScaler) para garantizar que todas las variables tuvieran la misma escala y, por ende, la misma importancia en el análisis.

El resultado del PCA se presenta en la Figura 4, que muestra la varianza explicada acumulada en función del número de componentes principales. A partir del gráfico, se observa que los primeros seis componentes principales son suficientes para explicar el 95% de la varianza en los datos. Esto significa que la mayoría de la información contenida en las variables originales puede ser representada de manera efectiva utilizando únicamente estos seis componentes, lo que permite simplificar el modelo sin perder precisión significativa.

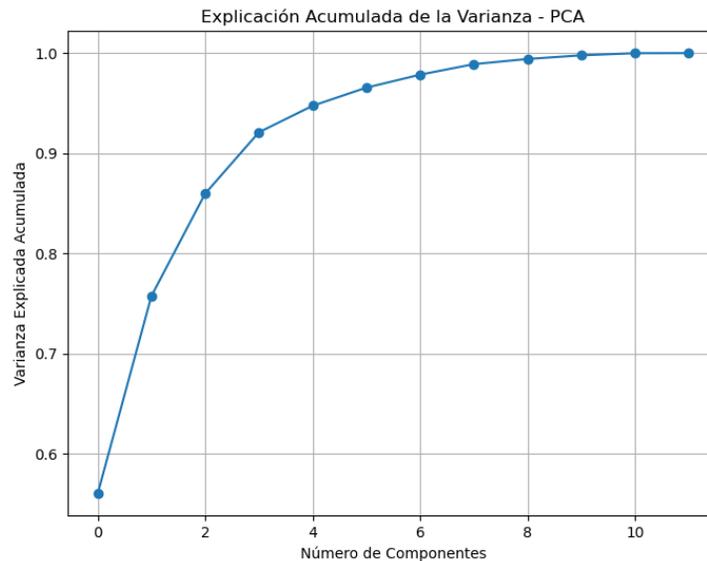


Figura 4: Explicación acumulada de la varianza (PCA)

Esta reducción de dimensionalidad no solo facilita el análisis, sino que también contribuye a evitar problemas como la colinealidad entre variables y el sobreajuste, optimizando así la eficiencia y robustez del modelo predictivo.

4.2.2. Tipos de análisis de correlaciones

El análisis de correlación es una herramienta esencial para evaluar la relación entre las variables recopiladas. En este estudio, se llevarán a cabo tres tipos principales de análisis de correlación: Pearson, Spearman y Kendall. Cada uno de ellos ofrece un enfoque diferente para identificar y medir la relación entre pares de variables, dependiendo de las características de los datos y del tipo de relación que se busca detectar. A continuación, se detalla cada tipo de análisis y su aplicabilidad.

El coeficiente de correlación de Pearson se utiliza para detectar relaciones lineales entre variables continuas. Este método es particularmente sensible a la presencia de valores atípicos, lo que puede influir significativamente en los resultados obtenidos. Sin embargo, cuando se cumplen los supuestos de linealidad y normalidad, Pearson es una herramienta robusta que permite cuantificar con precisión la fuerza y la dirección de la relación entre dos variables.

Por otro lado, el coeficiente de correlación de Spearman es adecuado para analizar relaciones monótonas entre variables continuas u ordinales. A diferencia de Pearson, Spearman se basa en los rangos de los datos en lugar de sus valores absolutos, lo que lo hace menos sensible a los valores atípicos. Este enfoque permite detectar relaciones que no necesariamente son lineales, pero que mantienen una tendencia consistente a medida que cambian los valores de las variables.

Finalmente, el coeficiente de correlación de Kendall se enfoca en medir la concordancia entre pares de datos, identificando relaciones monótonas. Este método, al igual que Spearman, es aplicable a datos continuos u ordinales, pero es aún menos sensible a los valores atípicos. Su interpretación se basa en la concordancia o discordancia entre pares de observaciones, lo que lo convierte en una herramienta útil para conjuntos de datos con características complejas o distribuciones no estándar.

4.2.3. Análisis de correlaciones con Pearson

La matriz de correlación de Pearson, presentada en la Figura 5, proporciona una visión integral de las relaciones lineales entre las variables consideradas en este estudio. Este análisis permitió

identificar redundancias y relaciones clave entre las variables, facilitando la selección de aquellas más relevantes para el modelo predictivo.

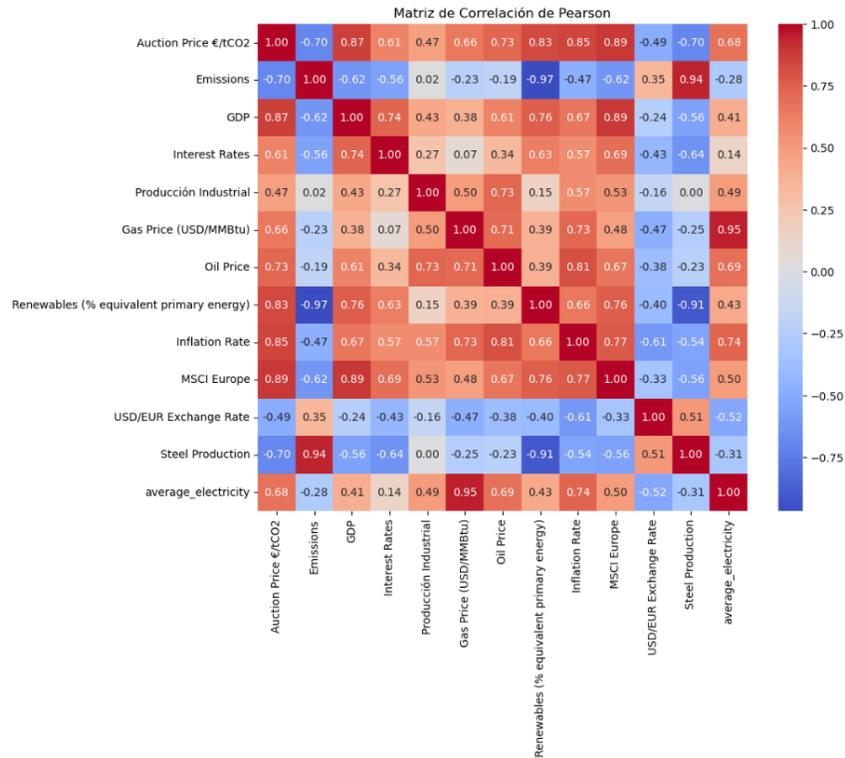


Figura 5: Matriz de correlación de Pearson

A partir de los coeficientes de correlación, se ha llevado a cabo un proceso de selección basado en criterios de redundancia e importancia contextual. Las variables seleccionadas para su inclusión en el modelo final son las siguientes:

1. **GDP:** Esta variable representa el crecimiento económico general y muestra correlaciones moderadas con otras variables, capturando aspectos amplios de la economía sin redundar excesivamente.
2. **Steel Production:** Aunque está altamente correlacionada con Emissions (~0.94), aporta información adicional relacionada con la actividad industrial intensiva en emisiones.
3. **Renewables (% equivalent primary energy):** Es altamente anticorrelacionada con Emissions (~-0.97) y resulta clave para capturar las dinámicas asociadas a la transición energética.

4. **Gas Price:** Representa el mercado energético con menos redundancia en comparación con otras variables, como Oil Price (~ 0.73).
5. **MSCI Europe:** Esta variable mide el desempeño del mercado financiero europeo, añadiendo una perspectiva única y complementaria al análisis.
6. **USD/EUR Exchange Rate:** Aunque sus correlaciones con otras variables son moderadas, ofrece un valor financiero distintivo y no presenta redundancia significativa.

Por otro lado, las siguientes variables fueron eliminadas debido a su alta correlación con otras o su carácter redundante:

1. **Producción Industrial:** Presenta una correlación alta con GDP (~ 0.74), lo que la hace redundante para el análisis.
2. **Emissions:** Aunque relevante, su fuerte correlación con Steel Production (~ 0.94) implica que no aporta información adicional significativa.
3. **Oil Price:** Más redundante que Gas Price, con elevadas correlaciones con múltiples variables.
4. **Inflation Rate:** Aunque relevante desde una perspectiva económica, su correlación con GDP (~ 0.85) y otras variables la hace menos distintiva.
5. **average_electricity:** Altamente correlacionada con Gas Price (~ 0.95), lo que la convierte en una variable redundante.

4.2.4. Análisis de correlaciones con Spearman

El análisis de correlaciones utilizando el coeficiente de Spearman, representado en la Figura 6, proporciona una perspectiva adicional sobre las relaciones monótonas entre las variables. A diferencia de Pearson, Spearman evalúa relaciones basadas en los rangos de las variables, lo que lo hace menos sensible a valores atípicos y, por tanto, particularmente útil en este contexto.

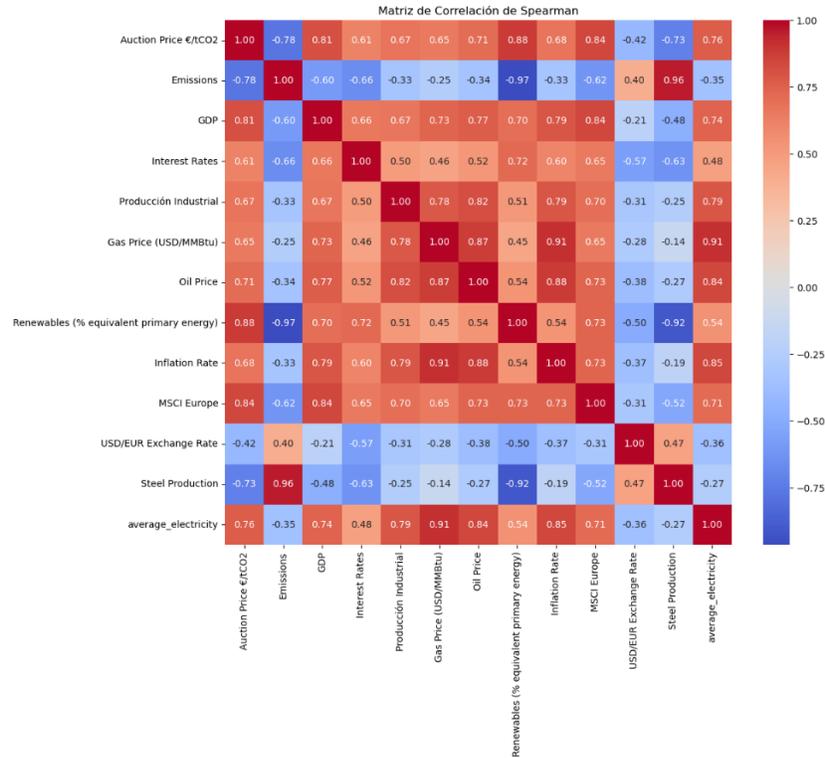


Figura 6: Matriz de correlación de Spearman

Con base en este análisis, se realizó una selección de variables considerando criterios de redundancia e importancia para el modelo predictivo. Las variables seleccionadas son las siguientes:

1. **GDP:** Esta variable refleja el crecimiento económico general y presenta correlaciones moderadas con otras variables sin llegar a ser redundante, lo que la hace fundamental para capturar dinámicas económicas amplias.
2. **Steel Production:** Aunque altamente correlacionada con Emissions (~ 0.96), es más representativa de la actividad industrial y aporta un enfoque más específico relacionado con sectores clave en las emisiones.
3. **Renewables (% equivalent primary energy):** Altamente anticorrelacionada con Emissions (~ -0.97), esta variable es esencial para reflejar la transición hacia fuentes de energía más limpias y sostenibles.
4. **Gas Price:** Más eficiente y menos redundante que Oil Price (~ 0.78), esta variable captura de manera efectiva las dinámicas del mercado energético.

5. **MSCI Europe:** Esta variable proporciona una perspectiva única sobre el desempeño financiero europeo y mantiene correlaciones manejables con otras variables.
6. **USD/EUR Exchange Rate:** Contribuye al análisis desde una dimensión financiera, sin mostrar redundancias significativas con otras variables seleccionadas.

Por otro lado, las siguientes variables fueron eliminadas debido a su alta correlación con otras o su menor contribución al modelo:

1. **Producción Industrial:** Presenta una correlación alta con GDP (~ 0.67), lo que la hace redundante en términos de información económica general.
2. **Emissions:** Aunque relevante, su correlación extremadamente alta con Steel Production (~ 0.96) implica que no añade información adicional significativa al modelo.
3. **Oil Price:** Más redundante que Gas Price, con múltiples correlaciones elevadas que la hacen prescindible.
4. **Inflation Rate:** Aunque importante desde una perspectiva económica, su alta correlación con GDP (~ 0.79) y otras variables la hace redundante.
5. **average_electricity:** Altamente correlacionada con Gas Price (~ 0.95), esta variable no aporta valor adicional significativo al modelo.

4.2.5. Análisis de correlaciones con Kendall

El análisis de correlaciones mediante el coeficiente de Kendall, representado en la Figura 7, permite identificar relaciones monótonas entre las variables recopiladas. Este método, basado en la concordancia y discordancia entre pares de observaciones, es particularmente útil en situaciones donde los datos contienen valores atípicos o presentan distribuciones no lineales. La menor sensibilidad de Kendall a valores extremos proporciona una visión más robusta de las interacciones entre variables.

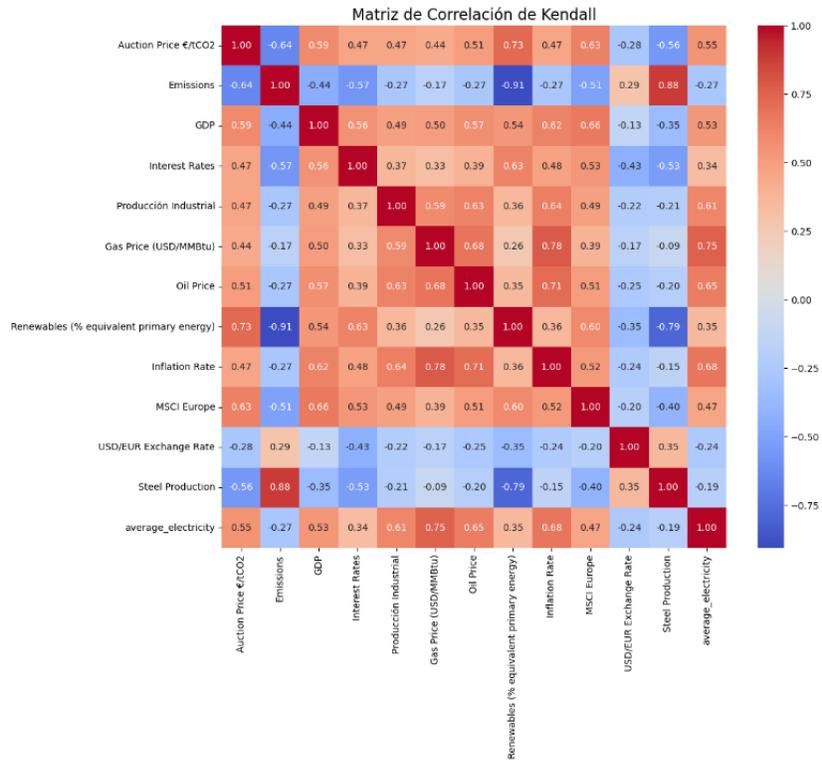


Figura 7: Matriz de correlación de Kendall

A partir de este análisis, se ha realizado una selección final de variables basándose en su relevancia y en la eliminación de redundancias. Las variables seleccionadas son las siguientes:

1. **GDP:** Representa el crecimiento económico global y mantiene correlaciones moderadas con otras variables. Su inclusión permite capturar dinámicas económicas amplias sin redundancia significativa.
2. **Steel Production:** Aunque está altamente correlacionada con Emissions (~ 0.88), ofrece una representación más específica de la actividad industrial intensiva en emisiones.
3. **Renewables (% equivalent primary energy):** Esta variable presenta una fuerte anticorrelación con Emissions (~ -0.91) y es esencial para reflejar las dinámicas asociadas a la transición energética hacia fuentes más limpias.
4. **Gas Price:** Más eficiente y menos redundante que Oil Price (~ 0.68), captura de manera efectiva las dinámicas del mercado energético.
5. **MSCI Europe:** Refleja la estabilidad y el desempeño del mercado financiero europeo, con correlaciones manejables respecto a otras variables.

6. **USD/EUR Exchange Rate:** Añade una dimensión financiera complementaria y no muestra redundancia significativa con las demás variables seleccionadas.

Por otro lado, se eliminaron las siguientes variables debido a su alta correlación con otras o a su carácter redundante en el contexto del modelo:

1. **Producción Industrial:** Su alta correlación con GDP (~ 0.49) la hace redundante al capturar aspectos similares de la economía.
2. **Emissions:** Aunque relevante, su fuerte correlación con Steel Production (~ 0.88) implica que no aporta información adicional única.
3. **Oil Price:** Más redundante que Gas Price, con correlaciones elevadas con múltiples variables.
4. **Inflation Rate:** Aunque importante, su correlación con GDP (~ 0.64) y otras variables económicas la convierte en prescindible.
5. **average_electricity:** Altamente correlacionada con Gas Price (~ 0.75), lo que la hace redundante en términos de información energética.

4.2.6. Evaluación de multicolinealidad

El Variance Inflation Factor (VIF) es una métrica utilizada para evaluar la multicolinealidad entre las variables independientes de un modelo. La multicolinealidad ocurre cuando una variable está altamente correlacionada con una o más variables del modelo, lo que puede afectar negativamente la estabilidad y la interpretabilidad del mismo.

El VIF mide en qué grado la varianza de un coeficiente de regresión se infla debido a la correlación entre variables. Valores de VIF por debajo de 5 suelen considerarse aceptables, mientras que valores por encima de 10 indican una alta multicolinealidad y sugieren redundancia en la información aportada por las variables. Un VIF elevado puede dificultar la identificación del impacto real de cada variable sobre la variable objetivo y provocar inestabilidad en los coeficientes estimados.

En la selección inicial, las seis variables seleccionadas tras el análisis de correlación mostraron los siguientes valores de VIF:

Variable	VIF
GDP	5.70
Steel Production	11.94
Renewables (% eqv. primary energy)	17.37
Gas Price	1.81
MSCI Europe	6.21
USD/EUR Exchange Rate	2.04

Tabla 4: Valores VIF de la selección inicial de variables

El elevado valor de VIF para Renewables (% equivalent primary energy) (17.37) indica una fuerte multicolinealidad con otras variables, como Steel Production (-0.91) y GDP (0.76). Esto la hace redundante y afecta la estabilidad del modelo.

Dado el alto VIF de Renewables (% equivalent primary energy), se tomó la decisión de reemplazarla por Oil Price, basándose en los siguientes factores:

- **Alta correlación con otras variables seleccionadas:** Renewables está altamente correlacionada con Steel Production (-0.91), GDP (0.76) y MSCI Europe (0.76), lo que reduce su capacidad para aportar información única al modelo.
- **Elevado VIF:** Su VIF de 17.37 indica una alta redundancia y un impacto negativo en la estabilidad del modelo.
- **Preferencia por variables menos redundantes:** Oil Price, aunque correlacionada con Gas Price (0.73), tiene una redundancia significativamente menor con las otras variables seleccionadas.
- **Mejora en la representación del mercado energético:** Oil Price complementa la información proporcionada por Gas Price, capturando dinámicas adicionales del mercado energético.

Tras reemplazar Renewables (% equivalent primary energy) por Oil Price, los nuevos valores de VIF son los siguientes:

Variable	VIF
----------	-----

GDP	5.52
Steel Production	2.22
Oil Price	3.36
Gas Price	2.31
MSCI Europe	5.94
MSCI Europe	1.79

Tabla 5: Valores VIF de la selección final de variables

Los nuevos resultados muestran una reducción significativa en la multicolinealidad:

- MSCI Europe tiene el VIF más alto (5.94), pero aún dentro de un rango aceptable.
- Oil Price y Gas Price presentan VIF bajos (3.36 y 2.31), lo que justifica su inclusión conjunta como representaciones complementarias del mercado energético.
- Steel Production y USD/EUR Exchange Rate tienen los VIF más bajos (2.22 y 1.79), indicando que son las variables menos redundantes del modelo.

El reemplazo de Renewables (% equivalent primary energy) por Oil Price mejoró la estabilidad y la interpretabilidad del modelo al reducir significativamente la multicolinealidad. Tras el ajuste basado en los valores de VIF, las seis variables seleccionadas para el modelo predictivo final son: GDP, Steel Production, Oil Price, Gas Price, MSCI Europe y USD/EUR Exchange Rate. Este conjunto de variables minimiza la redundancia, reduce la multicolinealidad y asegura que cada variable aporte información única y relevante para explicar las dinámicas que influyen en el precio de las emisiones de CO₂, proporcionando una base sólida para la construcción del modelo.

4.2.7. Conclusión del análisis de correlaciones

El análisis de correlación es un paso esencial para identificar las relaciones significativas entre las variables y la variable objetivo, en este caso, el precio de las emisiones de CO₂. Este proceso no solo permite seleccionar las variables más relevantes, sino que también contribuye a detectar redundancias o multicolinealidad que podrían comprometer la eficacia y la interpretabilidad del modelo predictivo. En este estudio, se aplicaron tres métodos de análisis de correlación (Pearson,

Spearman y Kendall) para evaluar desde relaciones lineales hasta monótonas. Además, este análisis se enriqueció con la evaluación del Variance Inflation Factor (VIF), que permitió controlar la multicolinealidad y asegurar que las variables seleccionadas aportaran información única al modelo.

Por otra parte, el análisis de componentes principales (PCA) destacó que, con solo seis variables, era posible explicar el 95% de la varianza de los datos. Este hallazgo validó que la selección de variables basada en correlación y VIF no solo era suficiente, sino también eficiente, al garantizar que el modelo capturara la mayor cantidad de información posible con el menor número de variables.

La Tabla 6 resume las variables seleccionadas para formar parte del modelo predictivo final, junto con sus coeficientes de correlación con la variable objetivo según los tres métodos utilizados. Estas variables fueron elegidas por su relevancia y capacidad para representar dinámicas clave en el comportamiento del precio de las emisiones de CO₂. Además, su selección tuvo en cuenta los valores de VIF, asegurando que cada una contribuyera de forma única al modelo, minimizando la redundancia y maximizando la estabilidad y la interpretabilidad del modelo.

Variable	Pearson	Spearman	Kendall
GDP	0.87	0.81	0.59
Steel Production	-0.7	-0.73	-0.56
Oil Price	0.73	0.71	0.51
Gas Price	0.66	0.65	0.44
MSCI Europe	0.89	0.84	0.63
USD/EUR Exchange Rate	-0.49	-0.42	-0.28

Tabla 6: Correlaciones de las variables seleccionadas

La Tabla 7 recoge las variables que, aunque presentaron correlaciones significativas con la variable objetivo, fueron excluidas del modelo tras un análisis detallado. Estas exclusiones se debieron a su alta redundancia, detectada tanto en las correlaciones como en los valores de VIF, o a su contribución limitada en comparación con las variables seleccionadas. Aunque algunas mostraron correlaciones relativamente altas, su inclusión habría introducido problemas de multicolinealidad

o complejidad innecesaria, lo que podría haber afectado negativamente la estabilidad y la generalización del modelo.

Variable	Pearson	Spearman	Kendall
Emissions	-0.7	-0.78	-0.64
Producción Industrial	0.47	0.67	0.47
Inflation Rate	0.85	0.68	0.47
Renewables (% equivalent energy)	0.83	0.88	0.73
average_electricity	0.68	0.76	0.55
Interest Rates	0.85	0.61	0.47

Tabla 7: Correlaciones de las variables eliminadas

La integración del análisis de correlación con los métodos de Pearson, Spearman y Kendall, junto con la evaluación del VIF y los resultados del PCA, permitió una selección óptima de variables para el modelo predictivo. Este enfoque permitió identificar un conjunto de seis variables que explican el 95% de la varianza en los datos, asegurando una representación adecuada de las dinámicas subyacentes. La exclusión de variables redundantes o poco relevantes optimizó la estabilidad, la interpretabilidad y el desempeño del modelo, sentando así una base sólida para predecir el precio de las emisiones de CO₂.

4.3. Análisis Descriptivo de las Variables

El análisis descriptivo de las variables es un paso fundamental en cualquier estudio de datos, ya que permite comprender las características principales de cada variable antes de avanzar hacia modelos más complejos. En esta sección, presentaremos estadísticas clave, como la media, la mediana y la desviación estándar, para ofrecer un panorama general del comportamiento de las variables. Estas métricas ayudarán a identificar tendencias, rangos y posibles valores atípicos.

Además, se incluirán visualizaciones individuales, como histogramas y diagramas de caja (boxplots), que facilitarán el análisis visual de las distribuciones y proporcionarán una

comprensión más profunda de las dinámicas de cada variable. Estas gráficas también son útiles para detectar patrones, valores extremos y posibles anomalías en los datos.

4.3.1. Producto Interior Bruto

El Producto Interior Bruto (PIB) mide el valor total de bienes y servicios producidos dentro de un país en un período específico. Es uno de los indicadores más importantes para evaluar el desempeño económico de un país y sirve como referencia para analizar tendencias económicas, comparar economías nacionales y evaluar políticas públicas. En este análisis, se realiza un estudio estadístico del PIB para identificar las principales métricas, como el promedio, la mediana y los cuartiles.

En el resultado del análisis descriptivo, se observan los siguientes puntos clave:

- **Promedio (mean):** El valor promedio del PIB es aproximadamente 16,357 unidades.
- **Desviación estándar (std):** La variación promedio del PIB respecto a su media es de 1,146 unidades.
- **Valores extremos:** El PIB mínimo registrado es de 14,765 unidades, mientras que el máximo es de 18,349 unidades.
- **Distribución:** Los cuartiles muestran que el 50% de los valores del PIB se encuentran entre 15,381 y 17,315 unidades.

	GDP
count	1029.000000
mean	16357.000275
std	1146.598283
min	14765.880177
25%	15381.418584
50%	15981.445870
75%	17315.127119
max	18349.388451

Output 1: Análisis descriptivo de PIB

Se genera un gráfico para visualizar la evolución del PIB a lo largo del tiempo, lo cual permite identificar tendencias de crecimiento o contracción económica:

- Se observa un patrón escalonado que indica incrementos o ajustes periódicos en el PIB.

- En ciertos períodos, se pueden identificar crecimientos marcados, reflejando posibles expansiones económicas.
- Las disminuciones podrían estar asociadas con eventos económicos significativos, como crisis económicas o recesiones.

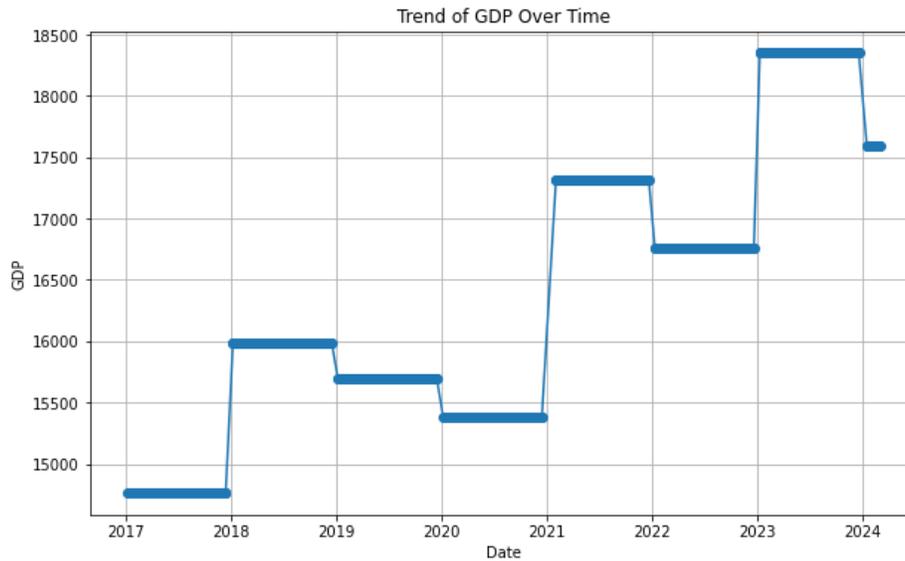


Figura 8: Gráfico de serie temporal de PIB

4.3.2. Producción de acero

La producción de acero es un indicador clave de la actividad industrial y económica, especialmente en sectores intensivos en emisiones de carbono. El acero es fundamental para múltiples industrias, desde la construcción hasta la fabricación de vehículos, y su producción refleja tanto la demanda económica como las políticas industriales vigentes. En este análisis, se realiza una descripción estadística detallada de la producción de acero para identificar sus principales características, incluyendo valores promedio, desviación estándar y rangos intercuartílicos.

El análisis descriptivo muestra los siguientes resultados clave para la producción de acero:

- **Promedio (mean):** El valor promedio de la producción de acero es aproximadamente 145.03 millones de toneladas.
- **Desviación estándar (std):** La variación promedio en la producción de acero respecto a su media es de 12.87 millones de toneladas.

- **Mínimo:** El valor más bajo registrado es de 126.19 millones de toneladas.
- **Máximo:** El valor más alto registrado es de 160.87 millones de toneladas.
- **Distribución intercuartílica:** el 50% de los valores se encuentran entre 132.22 millones (25%) y 160.07 millones (75%).

Steel Production	
count	1.029000e+03
mean	1.450343e+08
std	1.286557e+07
min	1.261854e+08
25%	1.322173e+08
50%	1.502444e+08
75%	1.600658e+08
max	1.608690e+08

Output 2: Análisis descriptivo de la producción de acero

Se genera un gráfico para visualizar la evolución de la producción de acero a lo largo del tiempo, lo que permite identificar tendencias clave, fluctuaciones y posibles patrones recurrentes.

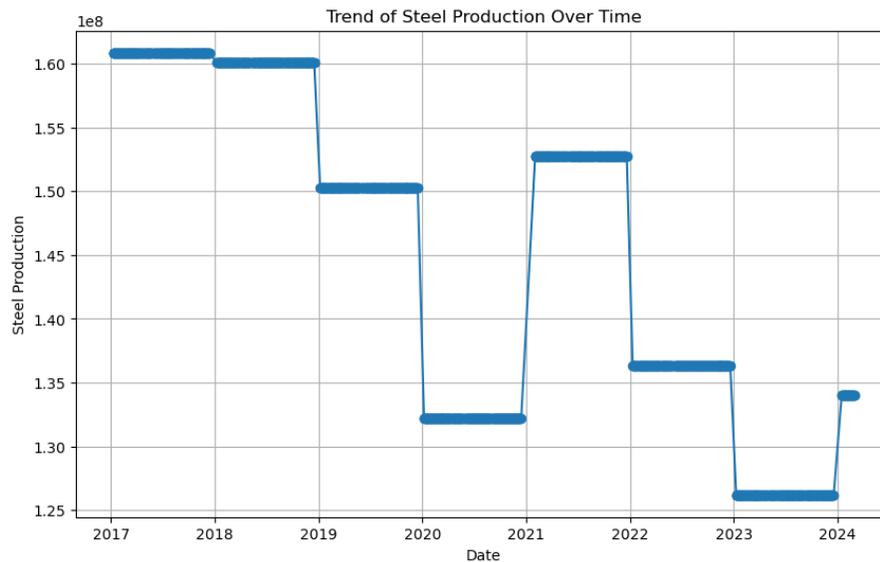


Figura 9: Gráfico de serie temporal de la producción de acero

El gráfico de tendencia temporal muestra los siguientes puntos destacados:

- **Tendencia general:** La producción de acero presenta un patrón escalonado, con variaciones significativas en ciertos períodos.

- **Descensos marcados:** Se observan disminuciones claras en varios puntos, posiblemente relacionadas con crisis económicas o ajustes en la demanda industrial.
- **Recuperación económica:** Algunos períodos reflejan un aumento notable, indicando fases de expansión económica o incrementos en la demanda de acero.
- **Estabilidad:** Durante ciertos intervalos, la producción se mantiene constante, lo que podría reflejar estabilidad en la industria o regulación del mercado.

4.3.3. Precio del petróleo

El precio del petróleo es un indicador clave que afecta a diversas áreas de la economía, incluyendo la producción de energía y la industria del transporte. Este precio también tiene una relación directa con los costos de producción de las emisiones de CO₂, ya que el petróleo es una fuente importante de energía que contribuye a las emisiones de gases de efecto invernadero. En este análisis, se exploran las características principales del precio del petróleo y su comportamiento a lo largo del tiempo, así como su distribución y los valores atípicos que podrían influir en el modelo predictivo.

El análisis descriptivo de los precios del petróleo muestra los siguientes resultados clave:

- **Promedio (mean):** El precio promedio del petróleo es de aproximadamente 70.11 unidades.
- **Desviación estándar (std):** La variabilidad en los precios del petróleo respecto a su media es de 19.87 unidades.
- **Mínimo:** El valor más bajo registrado es de 9.12 unidades.
- **Máximo:** El valor más alto registrado es de 133.18 unidades.
- **Distribución intercuartílica:** el 50% de los valores del precio del petróleo se encuentran entre 58.29 unidades (25%) y 82.34 unidades (75%).

	Oil Price
count	1029.000000
mean	70.108282
std	19.865670
min	9.120000
25%	58.290000
50%	69.650000
75%	82.340000
max	133.180000

Output 3: Análisis descriptivo del precio del petróleo

Para entender mejor la distribución del precio del petróleo, se visualiza mediante un histograma y un gráfico de caja. Estas visualizaciones permiten observar la frecuencia de aparición de ciertos rangos de precios y los posibles valores atípicos.

El histograma muestra una distribución bastante simétrica, con una concentración más alta de precios alrededor de las 60-80 unidades, lo que indica que la mayoría de los datos se encuentran dentro de este rango. Hay una ligera presencia de precios extremos, reflejados en las colas de la distribución.

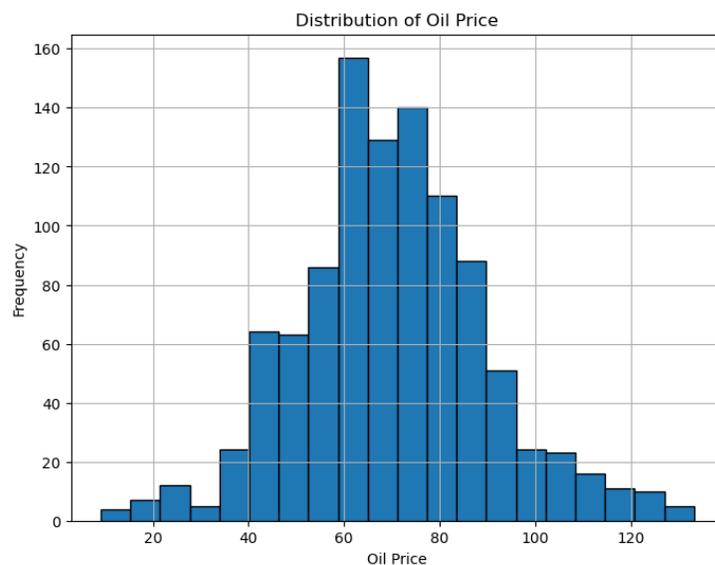


Figura 10: Histograma del precio del petróleo

El boxplot revela que el precio del petróleo tiene varios valores atípicos por encima de 100 unidades, lo que indica que durante ciertos períodos, los precios del petróleo han experimentado incrementos significativos.

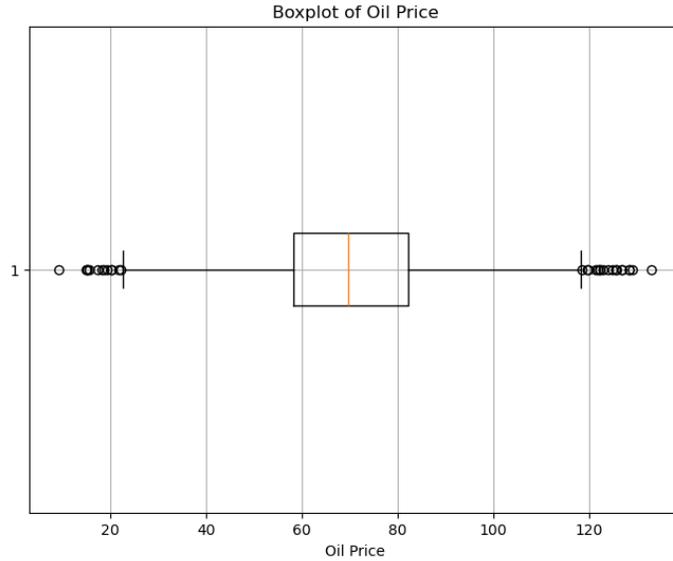


Figura 11: Boxplot del precio del petróleo

Se genera un gráfico de series temporales para analizar cómo ha evolucionado el precio del petróleo a lo largo del tiempo. Este gráfico muestra las fluctuaciones y tendencias significativas, lo cual es crucial para entender las dinámicas de oferta y demanda en el mercado global del petróleo.

El gráfico revela un patrón muy volátil en los precios del petróleo, con aumentos y caídas drásticas, especialmente a partir de 2020, lo que puede estar relacionado con eventos geopolíticos, crisis económicas o fluctuaciones en la producción y demanda global.

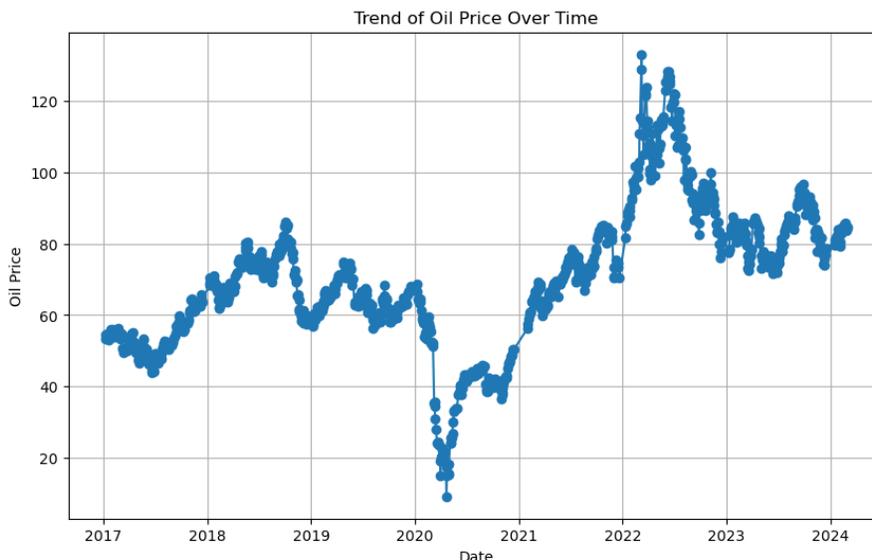


Figura 12: Gráfico de tendencia temporal del precio del petróleo

4.3.4. Precio del gas natural

El precio del gas natural es un factor clave en el mercado energético global, influyendo significativamente en los costos de producción y en la generación de emisiones de CO₂. Este precio afecta tanto a las decisiones de las empresas energéticas como a la dinámica de los mercados de carbono. En este análisis, se estudian las características principales del precio del gas natural, su distribución y su evolución temporal.

El análisis descriptivo del precio del gas natural muestra los siguientes resultados clave:

- **Promedio (mean):** El precio promedio del gas natural es de aproximadamente 12.58 USD/MMBtu.
- **Desviación estándar (std):** La variabilidad en los precios del gas natural respecto a su media es de 13.03 USD/MMBtu.
- **Mínimo:** El valor más bajo registrado es de 1.46 USD/MMBtu.
- **Máximo:** El valor más alto registrado es de 69.98 USD/MMBtu.
- **Distribución intercuartílica:** el 50% de los valores del precio del gas se encuentran entre 4.95 USD/MMBtu (25%) y 13.71 USD/MMBtu (75%).

	Gas Price (USD/MMBtu)
count	1029.000000
mean	12.576765
std	13.033938
min	1.462612
25%	4.953250
50%	7.504423
75%	13.714316
max	69.977239

Output 4: Análisis descriptivo del precio del gas natural

Para comprender la distribución del precio del gas natural, se generan un histograma y un gráfico de caja. Estas visualizaciones destacan la frecuencia de los precios y los valores atípicos.

El histograma muestra que los precios del gas tienen una distribución asimétrica positiva, con una alta concentración en los valores más bajos (entre 0 y 10 USD/MMBtu). Sin embargo, se observan precios más altos en las colas de la distribución, lo que sugiere períodos de incremento significativo.

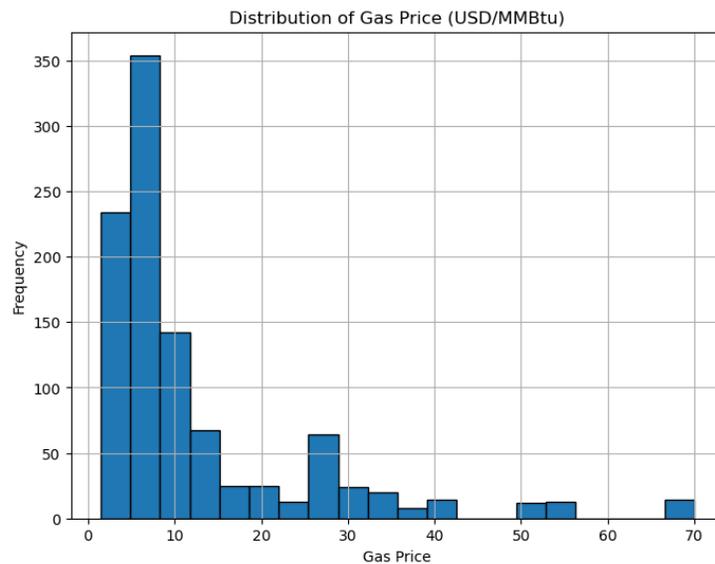


Figura 13: Histograma del precio del gas natural

El gráfico de caja revela la presencia de valores atípicos por encima de 30 USD/MMBtu, lo que indica picos inusuales en los precios durante ciertos períodos. Estos valores atípicos son importantes para entender eventos excepcionales en el mercado del gas.

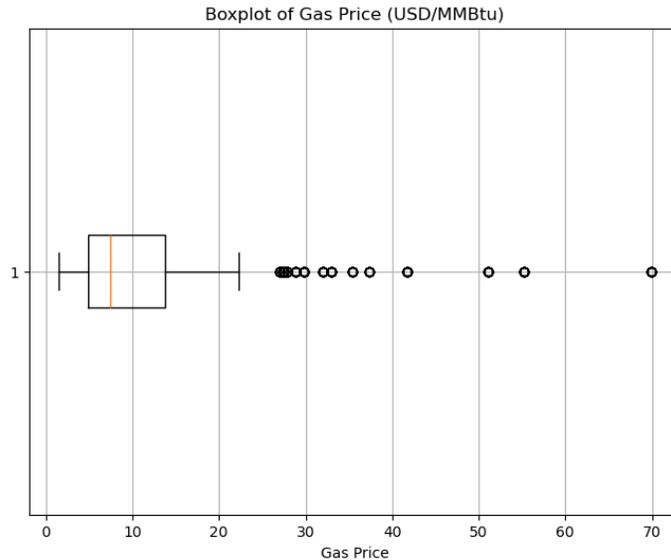


Figura 14: Boxplot del precio del gas natural

Se genera un gráfico de series temporales para analizar cómo ha evolucionado el precio del gas natural a lo largo del tiempo, lo que permite identificar patrones y fluctuaciones.

El gráfico de tendencia muestra las siguientes observaciones clave:

- **Estabilidad inicial:** Entre 2017 y 2021, los precios del gas se mantuvieron relativamente estables, con pequeñas fluctuaciones.
- **Incrementos marcados:** A partir de 2022, se observa un aumento abrupto en los precios, alcanzando un pico cercano a los 70 USD/MMBtu. Este período coincide con eventos significativos en los mercados energéticos globales.
- **Descenso posterior:** Tras alcanzar el máximo, los precios muestran una tendencia a la baja, estabilizándose en niveles más moderados en 2023 y 2024.

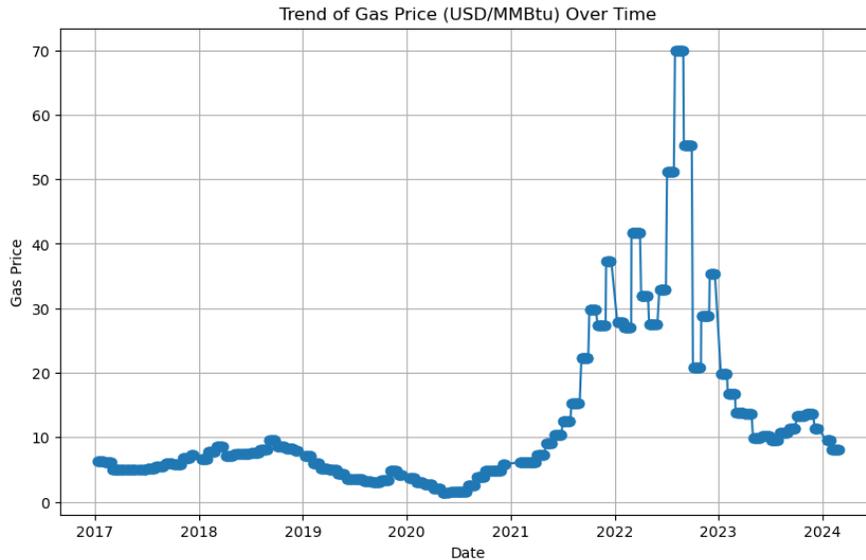


Figura 15: Gráfico de tendencia temporal del precio del gas natural

4.3.5. MSCI Europe

El índice MSCI Europe es un índice bursátil ampliamente utilizado que incluye empresas de gran y mediana capitalización en 15 mercados desarrollados de Europa. Este índice es un referente importante para los inversores que buscan medir el rendimiento del mercado de renta variable europeo. Al incluir compañías de sectores diversos, proporciona una visión integral de las condiciones económicas en la región. Es utilizado tanto como benchmark para comparar rendimientos como para construir estrategias de inversión basadas en los mercados europeos.

El análisis descriptivo de esta variable proporciona estadísticas clave como la media, los valores mínimos y máximos, los cuartiles y la desviación estándar. Estos datos nos ayudan a entender mejor la distribución de los valores del índice y su variabilidad.

El bloque de código realiza un análisis estadístico básico del índice MSCI Europe, mostrando estadísticas descriptivas como la media (promedio) de los valores, la mediana, la desviación estándar y los cuartiles.

En este output observamos:

- El índice tiene un valor promedio de 508,140.17, lo que representa el nivel medio durante el período analizado.

- La desviación estándar de 73,199.81 sugiere una variabilidad moderada en los valores del índice.
- El rango va desde un mínimo de 393,887.13 hasta un máximo de 666,916.69, destacando una amplia dispersión en los niveles del índice.
- El percentil 25 (primer cuartil) se encuentra en 448,215.38, mientras que el percentil 75 (tercer cuartil) se sitúa en 576,998.36, lo que indica que la mayoría de los valores están entre estos límites.

	MSCI Europe
count	1029.000000
mean	508140.169409
std	73199.813130
min	393887.132726
25%	448215.383578
50%	481186.974192
75%	576998.364156
max	666916.696867

Output 5: Análisis descriptivo de MSCI Europe

El histograma es una herramienta gráfica que permite observar la distribución de los valores del índice MSCI Europe. Es útil para identificar tendencias como la concentración de valores en ciertos rangos o posibles sesgos en la distribución.

El histograma muestra una distribución bimodal. Esto indica que los valores del índice tienden a agruparse en dos rangos distintos, reflejando posibles cambios significativos en el mercado durante el período analizado. Este comportamiento podría asociarse con eventos económicos o cambios estructurales en los mercados europeos.

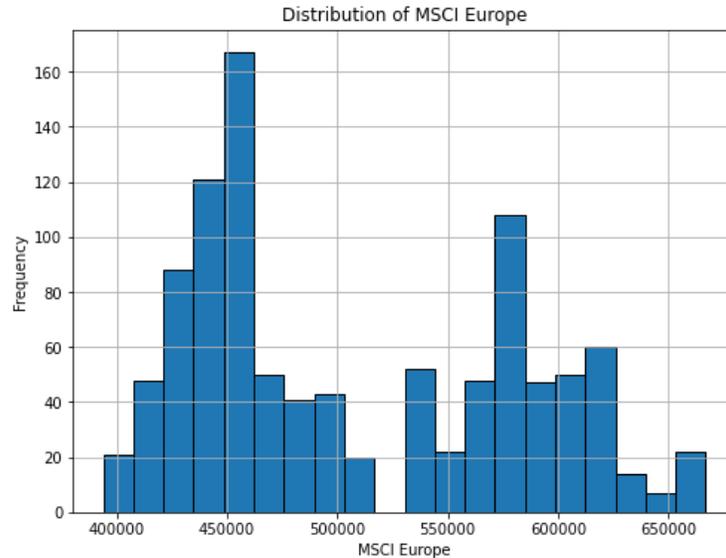


Figura 16: Histograma de MSCI Europe

El diagrama de caja, o boxplot, permite analizar de manera más precisa la dispersión de los datos, el rango intercuartil y la presencia de valores atípicos (outliers) en los niveles del índice MSCI Europe.

En el boxplot podemos observar que:

- La línea central del recuadro representa la mediana del índice, situada en 481,186.97, indicando que la mitad de los valores están por encima y la otra mitad por debajo de este nivel.
- El rango intercuartil (IQR), que abarca desde el primer cuartil (448,215.38) hasta el tercer cuartil (576,998.36), refleja la dispersión de los valores más comunes.
- Se detectan algunos valores atípicos en los extremos, indicando que en ciertos momentos el índice alcanzó niveles notablemente altos o bajos, posiblemente debido a fluctuaciones extremas del mercado.

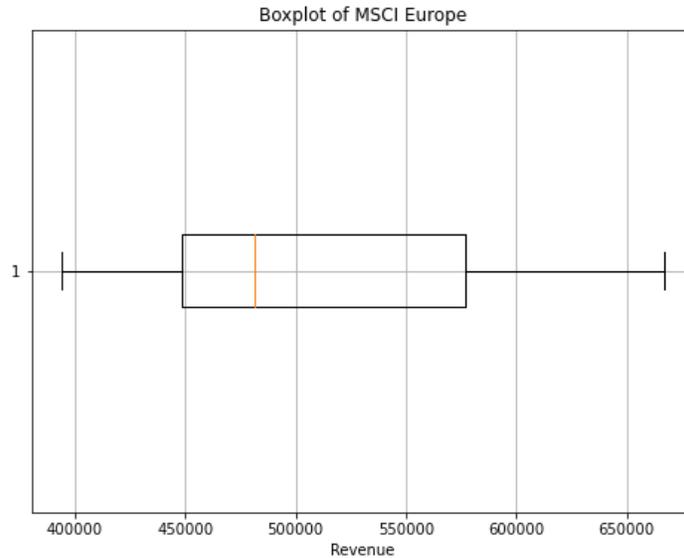


Figura 17: Diagrama de caja de MSCI Europe

El gráfico de tendencia temporal ilustra cómo ha evolucionado el índice MSCI Europe a lo largo del tiempo. Este análisis es crucial para identificar patrones de comportamiento y posibles puntos de inflexión en el mercado europeo.

El gráfico muestra un comportamiento general ascendente del índice MSCI Europe, lo que refleja un crecimiento consistente en el mercado de renta variable europeo durante el período analizado. Sin embargo, se observan algunos períodos de estabilidad y caídas abruptas que podrían estar relacionados con eventos económicos globales, tensiones geopolíticas o crisis específicas en la región. Estos cambios destacan la volatilidad inherente del mercado de acciones, especialmente en momentos de incertidumbre.

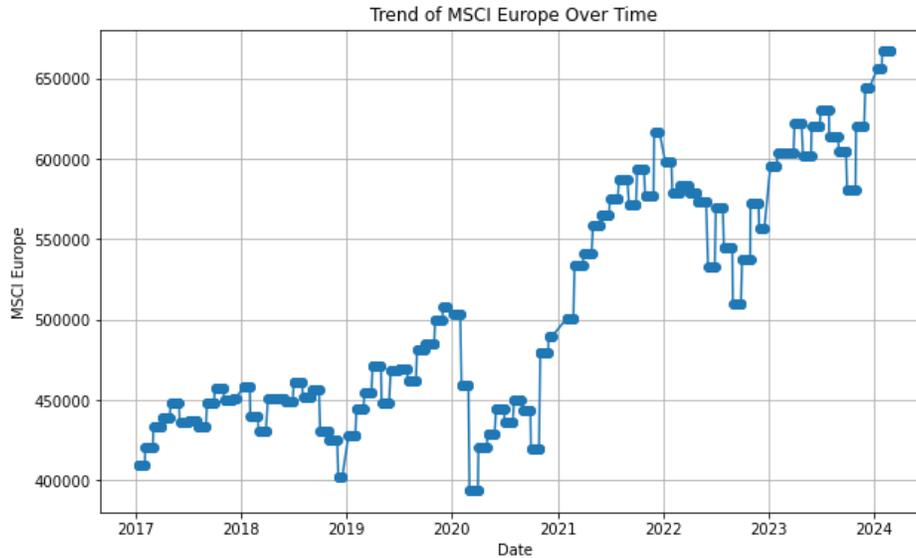


Figura 18: Gráfico de serie temporal de MSCI Europe

4.3.6. Tipo de cambio USD/EUR

El tipo de cambio USD/EUR es una variable financiera que indica cuántos euros (EUR) se necesitan para comprar un dólar estadounidense (USD). Este tipo de cambio fluctúa constantemente debido a varios factores, como las políticas monetarias, las tasas de interés, los indicadores económicos y los eventos globales. Es un indicador clave en la economía global y afecta directamente a las transacciones comerciales entre Estados Unidos y la Unión Europea, dos de las economías más grandes del mundo.

Este código proporciona un resumen estadístico básico del tipo de cambio USD/EUR. A continuación, se detallan métricas como el valor promedio, la desviación estándar, el valor mínimo, el valor máximo y los percentiles.

En el análisis descriptivo, observamos los siguientes resultados:

- **Media:** 1.124, lo que implica que, en promedio, 1 dólar estadounidense cuesta 1.124 euros.
- **Desviación estándar:** 0.057, lo que refleja cierta variabilidad moderada en el tipo de cambio.
- **Mínimo:** 0.964, lo que representa el valor más bajo registrado.
- **Máximo:** 1.249, que es el valor más alto registrado.

- Los percentiles indican la distribución de los valores, con el 50% de los datos (la mediana) en torno a 1.123 euros.

Esto sugiere una variabilidad significativa en los tipos de cambio durante el período analizado, aunque los valores se agrupan en torno a un rango estrecho.

	USD/EUR Exchange Rate
count	903.000000
mean	1.124382
std	0.057796
min	0.964400
25%	1.085400
50%	1.123000
75%	1.172900
max	1.249300

Output 6: Análisis descriptivo del tipo de cambio USD/EUR

El histograma visualiza la distribución de los tipos de cambio USD/EUR durante el período analizado. Este gráfico ayuda a identificar la frecuencia con la que los valores se concentran en diferentes intervalos.

En el histograma, se observa que la mayoría de los valores están agrupados en torno al rango de 1.10 a 1.15, lo que indica que este ha sido un rango común para el tipo de cambio. Sin embargo, también hay cierta dispersión hacia valores más bajos (alrededor de 0.96) y más altos (hasta 1.25), lo que resalta episodios de fortalecimiento o debilitamiento del dólar respecto al euro.

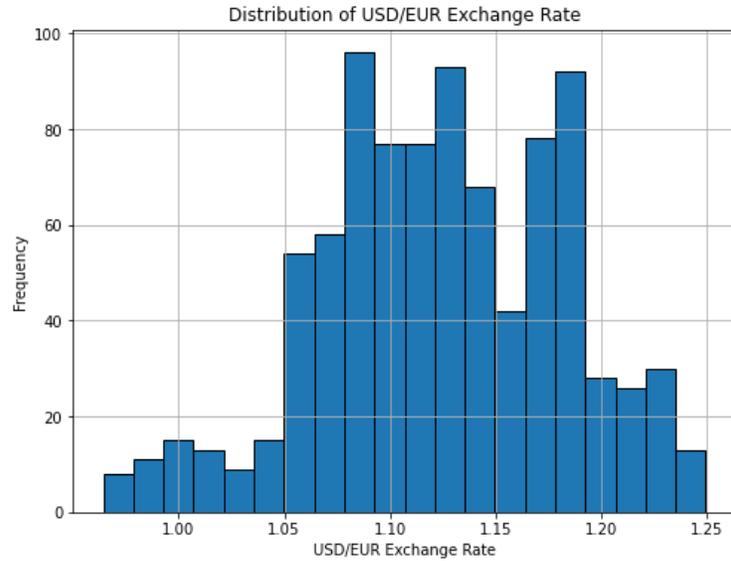


Figura 19: Histograma del tipo de cambio USD/EUR

El gráfico de serie temporal muestra claramente las fluctuaciones en el tipo de cambio USD/EUR a lo largo del tiempo. Se observan tendencias ascendentes y descendentes que corresponden a cambios económicos globales, como políticas monetarias de los bancos centrales, crisis económicas o tensiones comerciales entre la UE y Estados Unidos. La gráfica también revela períodos de relativa estabilidad intercalados con momentos de alta volatilidad, lo que destaca la naturaleza dinámica de los mercados de divisas.

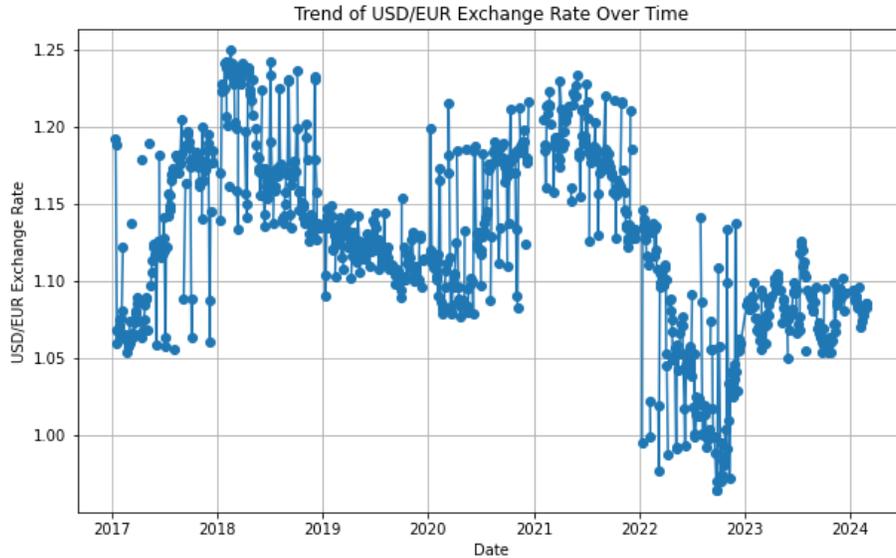


Figura 20: Gráfico de serie temporal del tipo de cambio USD/EUR

5. Análisis de Clustering

5.1. Propósito y Contexto del Clustering

El análisis de clustering es una técnica fundamental en la minería de datos y en la analítica avanzada, utilizada para segmentar un conjunto de datos en grupos homogéneos basados en patrones comunes. En este contexto, el objetivo principal es identificar subgrupos que compartan características similares, proporcionando una visión más detallada de las dinámicas subyacentes en el mercado de emisiones de carbono. Este enfoque permite no solo una mejor comprensión de los datos, sino también la inclusión de una nueva variable categórica en el modelo predictivo, enriqueciendo su capacidad para capturar relaciones complejas y no lineales.

En el caso específico de este análisis, el clustering es especialmente relevante debido a la naturaleza multifacética del precio de las emisiones de carbono. Este precio está influido por factores económicos, energéticos e industriales, tales como el crecimiento económico representado por el Producto Interior Bruto (GDP), los costos de energía medidos a través de los precios del petróleo y el gas, y la actividad industrial reflejada en la producción de acero. Estas variables, combinadas con indicadores financieros como el índice MSCI Europe y la tasa de cambio

USD/EUR, forman un panorama multidimensional que el clustering ayuda a descomponer en patrones más simples y manejables.

El propósito clave de este análisis es clasificar las observaciones en grupos que representen distintos contextos económicos y energéticos. Por ejemplo, un cluster podría reflejar períodos de alta producción industrial y bajos costos de energía, mientras que otro podría capturar períodos de altos precios de la energía y bajo desempeño económico. Estas diferencias entre clusters proporcionan una herramienta poderosa para identificar cómo las interacciones entre variables afectan al precio del carbono y permiten que el modelo predictivo considere estas dinámicas de manera explícita.

El algoritmo seleccionado para este análisis fue K-Means, ampliamente reconocido por su simplicidad y efectividad en tareas de segmentación. K-Means asigna cada observación a uno de los k clusters, minimizando la distancia entre los datos y el centroide del grupo. Este enfoque asegura que las observaciones dentro de un cluster sean lo más similares posible entre sí, mientras que las diferencias entre grupos sean maximizadas.

Además de mejorar la comprensión de los datos, el clustering tiene un propósito práctico directo: la inclusión de la clasificación resultante como una nueva variable categórica en el modelo predictivo. Esta variable aporta una dimensión adicional al análisis, capturando las diferencias entre grupos y ayudando a explicar las variaciones en el precio del carbono de una manera más robusta y precisa. En este sentido, el clustering no solo organiza y estructura los datos de manera significativa, sino que también refuerza la capacidad del modelo para capturar patrones complejos y mejorar sus predicciones.

5.2. Desarrollo y Resultados del Análisis

El análisis de clustering realizado en este trabajo tiene como objetivo segmentar las observaciones en grupos homogéneos para identificar patrones relevantes en las dinámicas de las variables seleccionadas. Estas dinámicas son clave para comprender los factores subyacentes que afectan al mercado de emisiones de carbono, permitiendo introducir un enfoque más granular en el modelo predictivo. En este análisis se utilizó el algoritmo de K-Means, que es ampliamente reconocido

por su eficacia en tareas de agrupamiento y su capacidad para trabajar con datos multidimensionales de manera eficiente.

5.2.1. Determinación del número de clusters: método del codo

El primer paso en el análisis fue determinar el número óptimo de clusters (k) que permitieran agrupar las observaciones de manera representativa sin sobreajustar el modelo. Para ello, se utilizó el método del codo, una técnica que evalúa la inercia, es decir, la suma de las distancias cuadradas entre los puntos y sus respectivos centroides, en función de diferentes valores de k . A medida que aumenta el número de clusters, la inercia disminuye, ya que cada grupo contiene menos variación interna. Sin embargo, después de cierto punto, el beneficio marginal de añadir más clusters disminuye significativamente, formando lo que se conoce como un “codo” en el gráfico.

En este análisis, el gráfico del método del codo, presentado en la Figura 21, muestra un punto de inflexión claro en $k=3$. Este resultado sugiere que tres clusters representan un equilibrio adecuado entre la simplicidad del modelo y la capacidad para capturar las diferencias entre las observaciones. Utilizar más de tres clusters habría implicado una complejidad innecesaria sin una mejora significativa en la calidad del agrupamiento, mientras que usar menos de tres habría pasado por alto variaciones importantes en los datos.

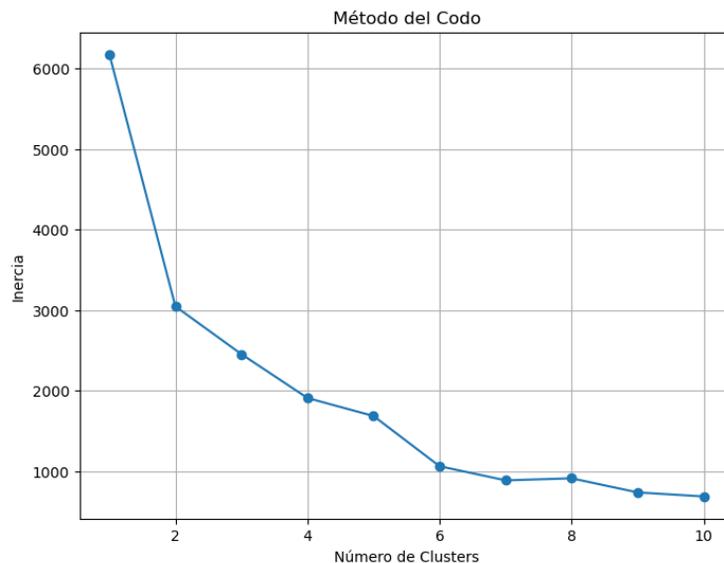


Figura 21: Método del codo

Con este resultado, se procedió a aplicar el algoritmo de K-Means con $k=3$, asegurando que cada observación se asignara a uno de los tres clusters. Este procedimiento permitió identificar patrones comunes dentro de cada grupo y diferencias significativas entre ellos.

5.2.2. Aplicación del algoritmo K-Means

El algoritmo de K-Means se basó en las variables seleccionadas previamente mediante análisis de correlaciones. Estas variables, como el GDP, la producción de acero, los precios del petróleo y el gas, el índice MSCI Europe y la tasa de cambio USD/EUR, reflejan factores económicos, industriales y energéticos que influyen directamente en el precio del carbono. Antes de aplicar el algoritmo, las variables fueron estandarizadas mediante el método StandardScaler, lo que garantizó que todas tuvieran el mismo peso en el proceso de agrupamiento. Esto fue esencial para evitar que características con valores numéricamente más grandes dominaran el análisis.

El algoritmo asignó cada observación a uno de los tres clusters, minimizando la distancia euclidiana entre los datos y el centroide de cada grupo. Este proceso se repitió iterativamente hasta que las posiciones de los centroides convergieron, asegurando que los grupos fueran lo más homogéneos posible internamente y lo más diferentes posible entre sí.

Tras completar el agrupamiento, se utilizó el análisis de componentes principales (PCA) para reducir las dimensiones de los datos y visualizar los clusters en un espacio bidimensional. La representación gráfica resultante, mostrada en la Figura 22, permite observar cómo las observaciones se agrupan de manera clara en tres clusters distintos. Cada grupo está bien separado de los demás, lo que confirma la efectividad del algoritmo en identificar diferencias significativas entre las observaciones.

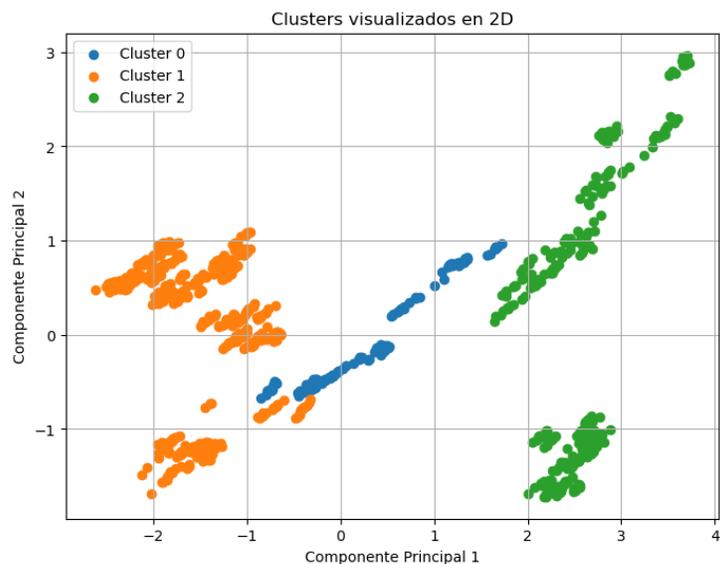


Figura 22: Clusters visualizados en 2D

5.2.3. Análisis de las características promedio de cada cluster

Para comprender mejor las diferencias entre los grupos, se calcularon las medias de las variables seleccionadas dentro de cada cluster. Los resultados, resumidos en la Tabla 1, reflejan las características clave de cada grupo y destacan las dinámicas económicas, energéticas e industriales que los diferencian.

Cluster	Date	Auction Price €/tCO ₂	GDP	Steel Production	Oil Price	Gas Price (USD/MMBtu)	MSCI Europe	USD/EUR Exchange Rate
0	2021-07-12 14:32:24.525547520	54.179197	17315.127119	1.527821e+08	72.156788	16.125658	564007.459399	1.181291
1	2019-01-02 16:13:22.787456512	17.748676	15460.290983	1.507287e+08	57.794983	5.284435	448762.826055	1.142224
2	2023-01-27 12:22:38.490566144	80.200943	17562.817123	1.314180e+08	91.451643	24.210722	591249.529022	1.067611

Figura 23: Valores promedio de las variables en cada cluster

El **Cluster 0** se caracteriza por un precio del carbono moderado, con un promedio de 54.18 €/tCO₂. Este grupo refleja un contexto económico robusto, con un GDP promedio alto de 17,315.13 € y una producción de acero elevada, alcanzando un promedio de 1.53×10^8 toneladas. Los precios del petróleo y del gas son moderados, situándose en 72.16 \$/barril y 16.13 \$/MMBtu, respectivamente.

El mercado financiero también muestra un desempeño sólido, con un índice MSCI Europe de 564,007 puntos. Este conjunto de características sugiere un entorno estable y equilibrado, donde los costos energéticos están controlados y la actividad económica es fuerte.

En el caso del **Cluster 1**, los precios del carbono son significativamente más bajos, con un promedio de 17.75 €/tCO₂. Este grupo se asocia con un GDP promedio más bajo de 15,460.29 € y precios de la energía notablemente reducidos, con un precio del petróleo de 57.79 \$/barril y un precio del gas de 5.28 \$/MMBtu. Aunque la producción de acero es alta, alcanzando un promedio de 1.51×10⁸ toneladas, el desempeño financiero es más débil, con un índice MSCI Europe de 448,762 puntos. Este cluster podría representar períodos de menor dinamismo económico global, combinados con menores costos de energía y una menor presión en los mercados de carbono.

Por último, el **Cluster 2** incluye los períodos con los precios del carbono más altos, con un promedio de 80.20 €/tCO₂. Este grupo se asocia con altos costos energéticos, reflejados en un precio promedio del petróleo de 91.45 \$/barril y un precio del gas de 24.21 \$/MMBtu. Aunque el GDP promedio es alto, situándose en 17,562.82 €, la producción de acero es más baja que en los otros clusters, con un promedio de 1.31×10⁸ toneladas. Este cluster refleja un contexto de mayor presión económica y regulatoria, posiblemente impulsado por altos costos de energía y un mercado financiero más robusto, con un índice MSCI Europe de 591,249 puntos.

5.2.4. Integración de la clasificación en el modelo predictivo

Tras completar el análisis, se añadió la clasificación en clusters como una nueva variable categórica al conjunto de datos original. Esta variable captura diferencias clave entre los grupos y enriquece el modelo predictivo al incorporar patrones y dinámicas complejas que influyen en el precio del carbono. Al segmentar las observaciones en función de sus características económicas, energéticas e industriales, esta variable permite que el modelo capte interacciones no lineales y mejore su precisión al explicar variaciones en el mercado de emisiones de carbono.

6. Modelos de Predicción

El núcleo de este trabajo radica en el desarrollo de modelos de predicción que permitan estimar el precio del carbono en el mercado EU ETS (Sistema de Comercio de Emisiones de la Unión Europea). Este mercado, uno de los pilares de las políticas europeas para combatir el cambio climático, fija un precio al carbono mediante un sistema de subastas y derechos de emisión, incentivando así la reducción de emisiones de gases de efecto invernadero. Debido a su alta volatilidad y a la influencia de múltiples factores económicos, regulatorios y energéticos, predecir su precio representa un desafío técnico significativo, pero a la vez, una tarea de gran relevancia para los sectores industriales y los responsables de políticas públicas.

El objetivo de este apartado es implementar y evaluar diferentes enfoques predictivos que abarquen desde modelos tradicionales de regresión hasta técnicas avanzadas basadas en aprendizaje automático y redes neuronales. Cada modelo tiene sus propias características, fortalezas y limitaciones, lo que permite comparar sus desempeños y explorar cuál es más adecuado para abordar la complejidad de los datos analizados. En particular, se busca no solo identificar un modelo que ofrezca un alto grado de precisión en las predicciones, sino también analizar su robustez, capacidad de generalización y viabilidad para aplicaciones prácticas en el contexto del mercado EU ETS.

Para lograr este objetivo, se seleccionaron modelos representativos de diferentes paradigmas en la predicción de datos. Se empieza con enfoques básicos como la regresión lineal y polinómica, que son útiles para establecer puntos de referencia y evaluar la linealidad de las relaciones entre las variables predictoras y el precio del carbono. Aunque estos modelos son sencillos y fácilmente interpretables, presentan limitaciones cuando los datos contienen relaciones no lineales o dependencias temporales.

A medida que se avanza, se exploran modelos de regresión regularizada, como Ridge y Lasso, que introducen penalizaciones en sus coeficientes para manejar problemas de sobreajuste y multicolinealidad. Estos métodos son especialmente relevantes en contextos donde las variables predictoras están altamente correlacionadas, lo que ocurre frecuentemente en conjuntos de datos relacionados con mercados financieros y energéticos.

Más allá de las regresiones, el análisis incorpora modelos no lineales basados en árboles de decisión y sus extensiones, como Random Forest, Gradient Boosting, y XGBoost. Estos modelos son particularmente eficaces en la captura de patrones complejos y no lineales, y su capacidad para manejar grandes cantidades de datos los convierte en herramientas fundamentales en aplicaciones modernas de aprendizaje automático. Su evaluación en este análisis no solo busca determinar su precisión, sino también comprender cómo pueden aprovechar las interacciones entre las variables para mejorar la predicción.

Por último, y como parte central del análisis, se implementan modelos basados en redes neuronales. Dentro de esta categoría, se incluyen el Perceptrón Multicapa (MLP) y las Redes Neuronales Recurrentes (LSTM). Estas últimas son especialmente relevantes debido a su capacidad para procesar series temporales, identificando dependencias a largo plazo que suelen ser cruciales en datos cronológicos como los del precio del carbono. Además, los modelos LSTM son adecuados para manejar la alta volatilidad y la naturaleza estacional del mercado EU ETS.

Cada modelo se entrena y evalúa utilizando un conjunto de datos que incluye variables económicas, regulatorias y energéticas relevantes para el precio del carbono. Para garantizar un análisis justo y sistemático, se adoptan métricas estándar de evaluación como el error cuadrático medio raíz (RMSE) y el coeficiente de determinación (R^2). El RMSE mide la magnitud promedio del error en las predicciones en las mismas unidades que los datos originales, mientras que el R^2 evalúa qué tan bien el modelo explica la variabilidad de los datos. Estas métricas, junto con gráficos comparativos de predicciones frente a valores reales, permiten identificar fortalezas y debilidades en cada enfoque.

Adicionalmente, para varios de los modelos se realiza una optimización de hiperparámetros, lo que implica ajustar configuraciones clave, como los valores de regularización en Ridge y Lasso o la profundidad de los árboles en Random Forest y Gradient Boosting. Este proceso asegura que los modelos operen bajo las mejores condiciones posibles, maximizando su rendimiento.

La importancia de desarrollar modelos precisos para predecir el precio del carbono en el mercado EU ETS no puede subestimarse. Un modelo confiable no solo tiene aplicaciones directas en la planificación estratégica de industrias intensivas en emisiones, sino que también ofrece herramientas valiosas para los responsables de diseñar políticas climáticas. La capacidad de

anticipar cambios en el precio del carbono puede facilitar la toma de decisiones informadas, tanto en la asignación de recursos como en la planificación de inversiones sostenibles.

Además, desde una perspectiva técnica, este análisis ofrece una visión integral de cómo diferentes enfoques predictivos pueden abordar un problema complejo como este. Por ejemplo, mientras que los modelos lineales son fácilmente interpretables y ofrecen una base sólida para entender relaciones generales, los enfoques no lineales y basados en redes neuronales son indispensables para capturar patrones complejos que pueden pasar desapercibidos en métodos más simples.

En los próximos apartados, se detallará la implementación de cada uno de los modelos seleccionados, desde su configuración inicial hasta los resultados obtenidos tras el entrenamiento y la evaluación. Asimismo, se presentará una comparación detallada de los modelos, destacando cuál de ellos ofrece el mejor balance entre precisión, robustez y capacidad de generalización. Finalmente, se discutirá cómo los resultados pueden ser interpretados en el contexto del mercado EU ETS y qué implicaciones tienen para futuras investigaciones y aplicaciones prácticas.

6.1. Regresión Lineal Múltiple

La regresión lineal múltiple permite construir un modelo lineal en el que el valor de la variable dependiente o de respuesta (Y) se calcula en función de un conjunto de variables independientes, denominadas predictores (X_1, X_2, X_3 , etc.). Es una generalización de la regresión lineal simple, por lo que resulta esencial comprender primero este concepto básico. Los modelos de regresión múltiple pueden utilizarse tanto para predecir el valor de la variable dependiente como para analizar la influencia de los predictores sobre ella (aunque este análisis debe realizarse con cuidado para evitar confundir correlación con causalidad). La ecuación que rige estos modelos lineales múltiples es la siguiente:

$$Y_i = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni}) + e_i$$

- β_0 : es la ordenada en el origen, el valor de la variable dependiente Y cuando todos los predictores son cero.

- β_i : es el efecto promedio que tiene el incremento en una unidad de la variable predictora X_i sobre la variable dependiente Y , manteniéndose constantes el resto de variables. Se conocen como coeficientes parciales de regresión.
- e_i : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo.

Es fundamental considerar que el valor de cada coeficiente parcial de regresión está influenciado por las unidades de medida de la variable predictora correspondiente, por lo que su magnitud no refleja necesariamente la importancia relativa del predictor. Para evaluar el impacto de cada variable en el modelo, se utilizan los coeficientes parciales estandarizados. Estos se obtienen al estandarizar las variables predictoras, lo que implica restar su media y dividir las entre su desviación estándar, antes de ajustar el modelo.

El modelo de regresión lineal múltiple se evaluó utilizando dos métricas principales: el Root Mean Squared Error (RMSE) y el coeficiente de determinación (R^2). Estos indicadores ayudan a entender la calidad de las predicciones realizadas por el modelo.

```
Evaluación del modelo global de regresión lineal:
Root Mean Squared Error (RMSE): 19.7604
R2 Score: -33.7906
```

Output 7: Métricas modelo de regresión lineal múltiple

- **Root Mean Squared Error (RMSE):** El valor obtenido, 19.7604, sugiere que el modelo presenta un error considerable al estimar los valores del precio del carbono. Esta magnitud refleja que las predicciones realizadas por el modelo se desvían significativamente de los valores reales, indicando una baja capacidad predictiva en este **contexto**.
- **R² Score:** El coeficiente de determinación obtenido, -33.7906, revela que el modelo no solo falla en capturar la variabilidad de los datos, sino que su desempeño es considerablemente peor que el de un modelo trivial que simplemente predijera el promedio de los valores observados. Un R^2 negativo tan elevado indica que el modelo es incapaz de explicar la relación entre las variables predictoras y el precio del carbono.

En el gráfico de residuos frente a predicciones, se observa una dispersión clara que no sigue ningún patrón sistemático. Sin embargo, la mayoría de los residuos están alejados de cero, lo que refuerza la idea de un ajuste pobre por parte del modelo. Este comportamiento puede deberse a la alta

complejidad de los datos, que no se adecúan bien a la suposición de linealidad inherente a este tipo de modelos.

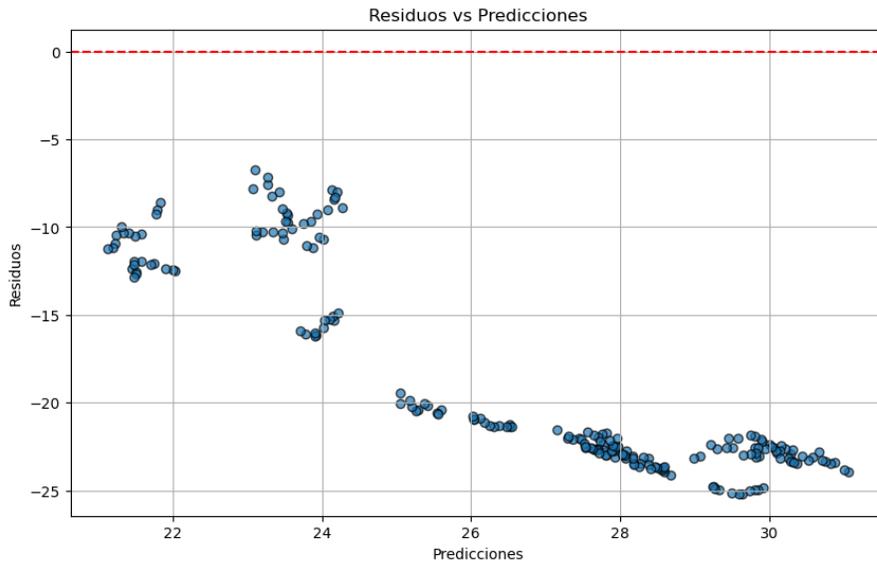


Figura 24: Relación entre los residuos y las predicciones del modelo de regresión lineal múltiple

La distribución de los residuos indica que los errores no están centrados alrededor de cero, lo que es otro indicativo de un mal ajuste del modelo. Además, la distribución no parece ser simétrica ni normal, lo que podría contradecir uno de los supuestos fundamentales de la regresión lineal.

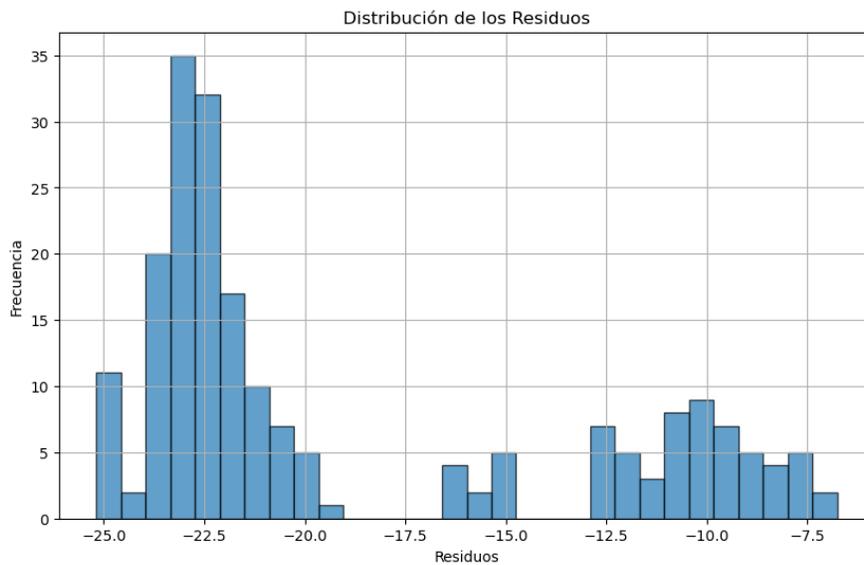


Figura 25: Histograma de la distribución de los residuos del modelo de regresión lineal múltiple

El gráfico compara los valores reales con los predichos, la dispersión de los puntos respecto a la línea ideal, donde las predicciones son iguales a los datos del test, es significativa. Los valores predichos no se alinean con los valores reales, lo que confirma la incapacidad del modelo para capturar patrones significativos en los datos.

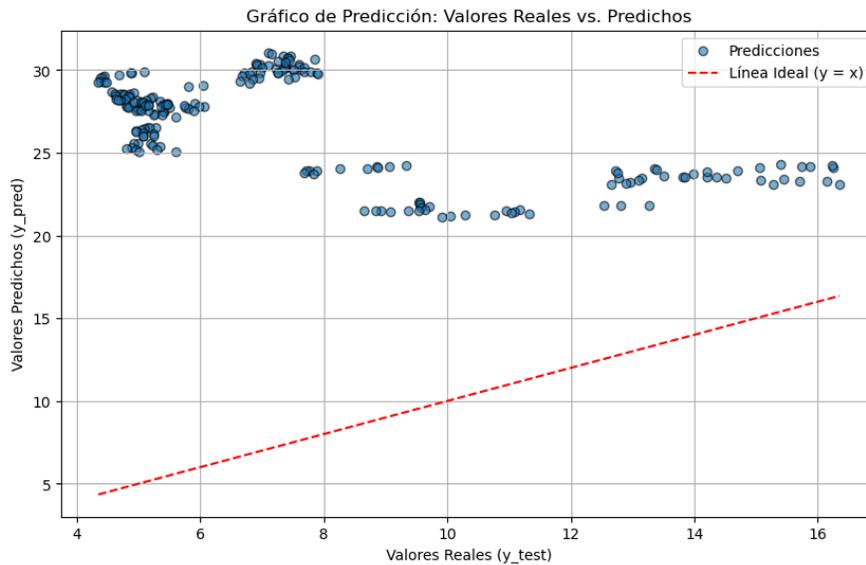


Figura 26: Comparación entre los valores reales y las predicciones del modelo de regresión lineal múltiple

El modelo de regresión lineal múltiple presenta un desempeño limitado en la predicción del precio del carbono en el mercado EU ETS, como lo reflejan su elevado RMSE (19.7604) y el R^2 negativo (-33.7906), lo que indica que no logra capturar adecuadamente la relación entre las variables predictoras y la variable objetivo.

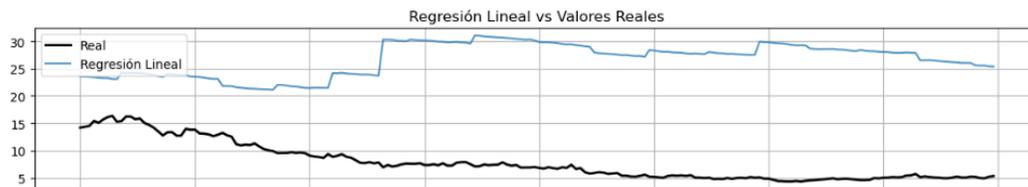


Figura 27: Comparación temporal de predicciones y valores reales del modelo de regresión lineal múltiple

El gráfico de comparación temporal (Figura 27) resalta cómo el modelo de regresión lineal múltiple falla en capturar adecuadamente las fluctuaciones del precio del carbono a lo largo del tiempo. Esto refuerza las limitaciones previamente descritas, evidenciando que la linealidad asumida por el modelo no es suficiente para reflejar la dinámica real de los datos. Por ello, resulta crucial explorar modelos más avanzados que sean capaces de captar relaciones más complejas.

6.2. Regresión Polinómica

La regresión polinómica extiende el concepto de la regresión lineal al incluir términos no lineales de las variables predictoras, permitiendo capturar relaciones más complejas entre las variables independientes y la dependiente. Este enfoque es particularmente útil cuando la relación entre las variables no sigue un patrón estrictamente lineal, pero aún puede representarse como una combinación de términos polinómicos. A diferencia de la regresión lineal simple o múltiple, donde los predictores son considerados en su forma original, en la regresión polinómica se incluyen potencias de estos predictores, creando términos como X_1^2 , X_1^3 , y así sucesivamente, hasta el grado especificado del polinomio. La ecuación básica de un modelo de regresión polinómica de grado d se expresa como:

$$Y_i = \beta_0 + \beta_1 X_1 i + \beta_2 X_1 i^2 + \beta_3 X_1 i^3 + \dots + \beta_d X_1 i^d + e_i$$

- **β_0 :** Es el término independiente que representa el valor de Y cuando todos los predictores son cero.
- **β_i :** Coeficientes asociados a cada término del polinomio, que indican el impacto promedio de los incrementos en los términos elevados de las variables predictoras sobre la variable dependiente.
- **e_i :** Residuo o error, es decir, la diferencia entre los valores observados y los predichos por el modelo.

Es importante destacar que, aunque este enfoque mejora la flexibilidad del modelo al capturar patrones no lineales, también incrementa el riesgo de sobreajuste si se utilizan grados demasiado altos para los polinomios. Por ello, es crucial balancear la complejidad del modelo con su capacidad para generalizar en datos no vistos.

Para encontrar el grado óptimo de la regresión polinómica, se evaluaron diferentes modelos variando el grado del polinomio y analizando las métricas de rendimiento, como el RMSE y R^2 .

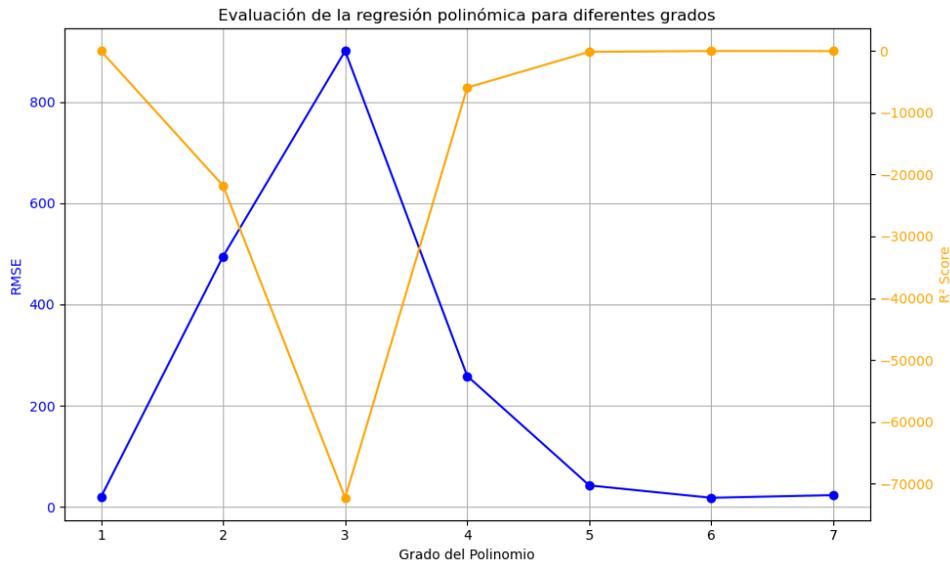


Figura 28: Evaluación del rendimiento de la regresión polinómica según el grado del polinomio

El análisis presentado destaca cómo varían estas métricas con respecto al grado del polinomio utilizado en la regresión. En este caso, el grado 6 se identifica como el más adecuado para el modelo. Esto se debe a que el RMSE alcanza su valor mínimo en este grado, lo que indica el menor error en las predicciones realizadas por el modelo. Simultáneamente, el coeficiente de determinación R^2 se estabiliza en un valor cercano a cero, lo que sugiere que el modelo es capaz de capturar las relaciones presentes en los datos sin introducir ruido innecesario.

Aunque grados superiores podrían mantener métricas similares, optar por el grado 6 evita añadir complejidad innecesaria al modelo, siguiendo el principio de parsimonia. La complejidad adicional podría aumentar el riesgo de sobreajuste y dificultar la interpretación del modelo.

Por otro lado, el comportamiento errático observado en grados intermedios, como el incremento pronunciado en el RMSE y la caída abrupta de R^2 en el grado 3, evidencia posibles problemas de inestabilidad numérica o sobreajuste, especialmente cuando se incrementa la complejidad del modelo sin una adecuada regularización. Esto refuerza aún más la elección del grado 6 como un

punto de equilibrio entre precisión y simplicidad, maximizando la capacidad del modelo para generalizar en datos no vistos y manteniendo su robustez.

El modelo de regresión polinómica de grado 6 fue evaluado utilizando las métricas de Root Mean Squared Error (RMSE) y el coeficiente de determinación (R^2), las cuales permiten analizar la calidad de las predicciones y el ajuste del modelo a los datos.

```
Evaluación del modelo final de regresión polinómica (grado 6):  
Root Mean Squared Error (RMSE): 18.0625  
R2 Score: -28.0689
```

Output 8: Métricas modelo de regresión polinómica (grado 6)

- **Root Mean Squared Error (RMSE):** El valor obtenido, 18.0625, indica un error promedio menor en comparación con el modelo de regresión lineal múltiple. Sin embargo, sigue reflejando un nivel significativo de desviación entre las predicciones y los valores reales, lo que limita la precisión del modelo en este contexto.
- **R² Score:** El coeficiente de determinación de -28.0689 señala que, aunque el modelo de grado 6 mejora levemente respecto a la regresión lineal múltiple, sigue siendo incapaz de explicar la variabilidad en los datos. El R^2 negativo sugiere que el modelo tiene un desempeño peor que un modelo trivial basado en el promedio de los valores observados.

En el gráfico de residuos frente a predicciones, se observa una distribución heterogénea de los residuos, con patrones evidentes de dispersión. Este comportamiento sugiere que el modelo tiene dificultades para capturar las relaciones no lineales de los datos, a pesar de incluir términos polinómicos. La mayoría de los residuos están alejados de cero, lo que refuerza la idea de que el modelo no logra un ajuste adecuado.

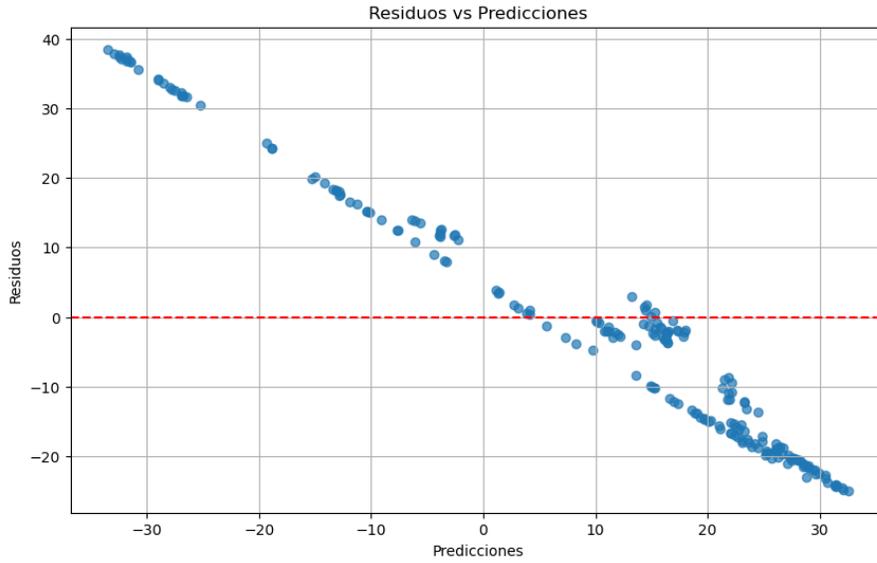


Figura 29: Relación entre los residuos y las predicciones del modelo de regresión polinómica

La distribución de los residuos presenta múltiples picos, lo que refleja que los errores no siguen una distribución normal ni simétrica. Este comportamiento contradice los supuestos básicos de regresión y apunta a posibles problemas de especificación o inestabilidad del modelo.

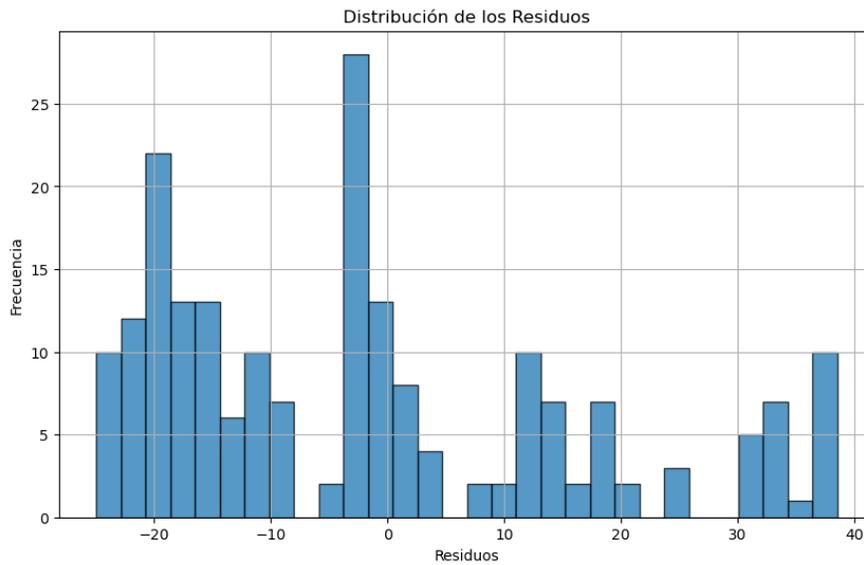


Figura 30: Histograma de la distribución de los residuos del modelo de regresión polinómica

El gráfico de valores reales frente a valores predichos muestra una dispersión significativa de los puntos respecto a la línea ideal ($y = x$). Esto evidencia que las predicciones del modelo no están alineadas con los valores reales y confirma que el modelo no es capaz de capturar patrones significativos en los datos.

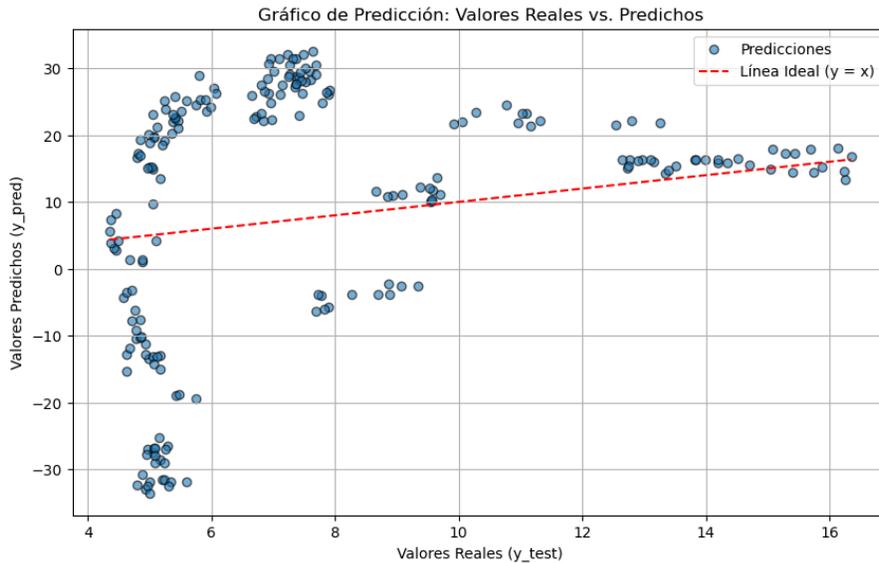


Figura 31: Comparación entre los valores reales y las predicciones del modelo de regresión polinómica

El modelo de regresión polinómica de grado 6 mejora ligeramente las métricas de evaluación en comparación con la regresión lineal múltiple, como lo demuestra el menor RMSE (18.0625). Sin embargo, el R^2 negativo (-28.0689) y el análisis de los gráficos indican que el modelo sigue siendo inadecuado para capturar la complejidad de los datos.



Figura 32: Comparación temporal entre los valores reales y las predicciones del modelo de regresión polinómica (grado 6)

El gráfico de comparación temporal (Figura 32) muestra cómo el modelo de regresión polinómica de grado 6 sigue sin captar las fluctuaciones más significativas del precio del carbono a lo largo

del tiempo. A pesar de incluir términos no lineales, el modelo presenta desviaciones notorias respecto a los valores reales, lo que refuerza las limitaciones identificadas en las métricas. Este resultado subraya la necesidad de adoptar enfoques más sofisticados para capturar las complejidades del conjunto de datos.

6.3. Lasso y Ridge

Los modelos Lasso y Ridge son métodos de regresión lineal regularizada que buscan mejorar la precisión predictiva y la estabilidad de los modelos lineales tradicionales, especialmente cuando los datos presentan multicolinealidad o cuando el número de predictores es mayor que el número de observaciones. Estos métodos introducen penalizaciones en los coeficientes de la regresión para evitar problemas de sobreajuste y simplificar el modelo. Ambos métodos pertenecen a la familia de regresiones lineales regularizadas, pero difieren en el tipo de penalización que aplican a los coeficientes:

La **regresión Lasso** agrega una penalización L1 a la función de costo. Este término penaliza la suma de los valores absolutos de los coeficientes, lo que puede llevar a que algunos coeficientes se reduzcan exactamente a cero, logrando una selección automática de variables. La fórmula de la función objetivo de Lasso es:

$$\text{Minimizar: } \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Donde:

Los términos son los mismos que en la fórmula de Ridge.

- λ controla la fuerza de la penalización:
 - Si $\lambda = 0$, Lasso se reduce a una regresión lineal tradicional.
 - Si λ es grande, más coeficientes se reducen a cero, eliminando efectivamente ciertas variables del modelo.

Lasso es particularmente útil cuando se trabaja con conjuntos de datos con muchas variables irrelevantes, ya que selecciona automáticamente las más relevantes.

La **regresión Ridge** agrega una penalización L2 a la función de costo de la regresión lineal. Este término penaliza la suma de los cuadrados de los coeficientes, reduciendo su magnitud y mitigando problemas de multicolinealidad. La fórmula de la función objetivo de Ridge es:

$$\text{Minimizar: } \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

Donde:

- y_i son los valores reales de la variable dependiente.
- \hat{y}_i son las predicciones del modelo.
- β_j son los coeficientes de las variables predictoras.
- $\lambda \geq 0$ es el hiperparámetro de regularización que controla la magnitud de la penalización:
 - Si $\lambda = 0$, Ridge se reduce a una regresión lineal tradicional.
 - Si λ es grande, los coeficientes tienden a acercarse a cero, lo que reduce la influencia de los predictores.

Ridge no fuerza los coeficientes a ser exactamente cero, sino que los reduce, manteniendo todas las variables en el modelo. Esto lo hace útil cuando se desea conservar todas las variables predictoras, pero con menor peso.

Aspecto	Lasso	Ridge
Penalización	$\lambda \sum \beta_j $	$\lambda \sum \beta_j^2$
Manejo de multicolinealidad	Bueno, pero depende de la selección de variables.	Bueno
Selección de variables	Puede reducir coeficientes a cero, eliminando variables.	No elimina variables
Uso principal	Cuando se busca seleccionar las variables más importantes.	Cuando todas las variables son relevantes, pero con menor peso.

Tabla 8: Comparación entre Lasso y Ridge

6.3.1. Regresión Lasso

El modelo Lasso introduce un término de regularización que, además de reducir el impacto de problemas como el sobreajuste y la multicolinealidad, selecciona automáticamente las variables más relevantes, ya que es capaz de reducir algunos coeficientes a exactamente cero. Este enfoque es especialmente útil en escenarios con un alto número de predictores, permitiendo identificar las variables más significativas para la predicción.

```
Evaluación del modelo Lasso:  
Root Mean Squared Error (RMSE): 5.8670  
R2 Score: -2.0670
```

Output 9: Métricas modelo Lasso

En su implementación inicial, el modelo Lasso obtuvo un RMSE de 5.8670 y un R^2 de -2.0670, lo que refleja un ajuste ligeramente mejor en comparación con algunos modelos anteriores. Sin embargo, el valor negativo del coeficiente de determinación R^2 continúa indicando que el modelo tiene un desempeño inferior al de un modelo trivial que predice simplemente el promedio de los valores observados.

Para incrementar la precisión de Lasso, se llevó a cabo una optimización del hiperparámetro α , que controla la magnitud de la penalización aplicada a los coeficientes. Dicho proceso consistió en explorar de forma sistemática distintos valores de α , ajustar el modelo en cada caso y calcular el RMSE resultante. De esta manera, se identificó qué valor concreto de α producía la menor tasa de error en los datos de validación o en un esquema de validación cruzada.

```
Mejor alpha para Lasso: 0.1453  
RMSE para Lasso optimizado: 5.7541  
R2 para Lasso optimizado: -1.9500
```

Output 10: Métricas modelo Lasso optimizado

A través de esta búsqueda, se determinó que el α óptimo es 0.1453. Con este ajuste, el modelo Lasso optimizado redujo el RMSE a 5.7541 y mejoró el R^2 a -1.9500. Aunque el coeficiente de determinación continúa siendo negativo, lo que indica que el modelo no supera a una predicción

basada en la media, la reducción del error cuadrático medio denota una ganancia en precisión frente a la versión inicial.

El gráfico de comparación temporal (Figura 33) muestra cómo el modelo Lasso optimizado sigue con mayor precisión las tendencias generales del precio del carbono, en comparación con el modelo no optimizado. Sin embargo, se evidencian limitaciones en su capacidad para capturar variaciones complejas, reflejando que, a pesar de la mejora observada, el modelo aún enfrenta desafíos en términos de ajuste y generalización.



Figura 33: Comparación temporal entre los valores reales y las predicciones del modelo Lasso optimizado

En la Figura 34, se analiza la relación entre los residuos y las predicciones realizadas por el modelo Lasso optimizado. Los residuos presentan una distribución más concentrada alrededor de cero en comparación con los modelos no optimizados, lo que indica un mejor ajuste global. Además, se observa una reducción en la magnitud de los valores extremos, lo que refleja que el modelo optimizado logra un ajuste más robusto.

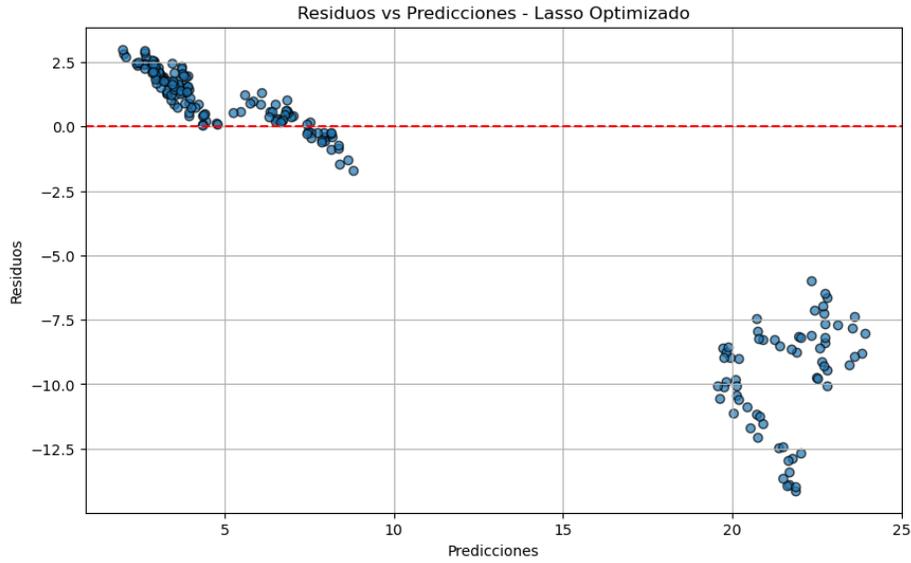


Figura 34: Relación entre los residuos y las predicciones del modelo Lasso optimizado

La Figura 35 muestra el histograma de la distribución de los residuos. En este gráfico, se aprecia que, aunque los residuos no siguen una distribución normal perfecta, la mayor parte de ellos se encuentra cerca de cero. Esto evidencia una mejora en el comportamiento del modelo tras la optimización del hiperparámetro α , con menos errores extremos en las predicciones.

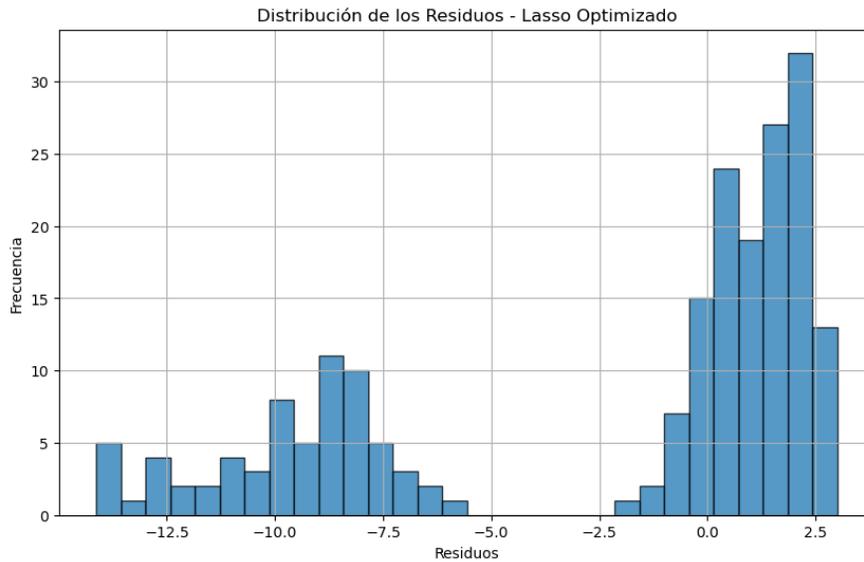


Figura 35: Histograma de la distribución de los residuos del modelo Lasso optimizado

En la Figura 36, se comparan los valores reales del conjunto de prueba con los valores predichos por el modelo Lasso optimizado. Aunque las predicciones no se alinean completamente con la línea ideal, se observa una disminución en la dispersión de los puntos en comparación con el modelo inicial y otros métodos evaluados. Este resultado sugiere que el modelo optimizado capta mejor las relaciones subyacentes en los datos.

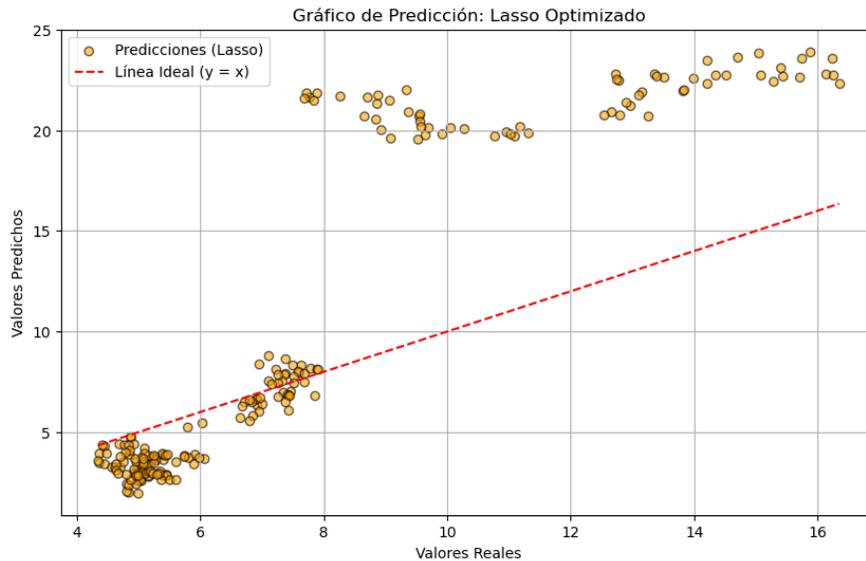


Figura 36: Comparación entre los valores reales y las predicciones del modelo Lasso optimizado

El modelo Lasso optimizado muestra un desempeño mejorado respecto a su versión inicial, como lo demuestran el menor RMSE (5.7541) y la mejora en el coeficiente R^2 (-1.9500). Sin embargo, el hecho de que R^2 permanezca negativo indica que el modelo sigue siendo incapaz de capturar adecuadamente la variabilidad de los datos en este contexto. Esto podría deberse a la complejidad inherente de los datos y a la posible necesidad de modelos no lineales o más sofisticados para lograr una precisión significativamente mayor.

6.3.2. Regresión Ridge

El modelo Ridge incorpora un término de regularización que reduce el impacto de problemas como la multicolinealidad y el sobreajuste en los datos. Este enfoque es especialmente útil en contextos con un alto número de variables predictoras altamente correlacionadas, como ocurre en los datos

analizados en este estudio. Para evaluar el desempeño del modelo, se emplearon dos métricas clave: el Root Mean Squared Error (RMSE) y el coeficiente de determinación (R^2).

```
Evaluación del modelo Ridge:  
Root Mean Squared Error (RMSE): 6.1887  
R2 Score: -2.4125
```

Output 11: Métricas del modelo Ridge

Con el fin de optimizar aún más el rendimiento del modelo, se llevó a cabo un proceso de búsqueda del valor óptimo del hiperparámetro α , que determina el nivel de regularización. Para ello, se utilizó una estrategia de minimización que consiste en variar sistemáticamente el valor de α , evaluar el rendimiento (medido por el RMSE) en cada caso y seleccionar aquel α que produce el RMSE más bajo. Este procedimiento implica entrenar el modelo repetidamente con distintos valores de α y calcular el error de predicción en un conjunto de validación o mediante validación cruzada, garantizando así la elección de la mejor regularización posible.

```
Mejor alpha para Ridge: 99.2515  
RMSE para Ridge optimizado: 5.5331  
R2 para Ridge optimizado: -1.7278
```

Output 12: Métricas del modelo Ridge optimizado

A raíz de este proceso de optimización, se encontró que el valor óptimo de α es 99.2515, lo que permitió reducir el RMSE a 5.5331 y mejorar ligeramente el R^2 a -1.7278. Estos resultados muestran que la introducción de regularización ayuda a reducir el error y estabilizar el modelo, aunque el coeficiente de determinación continúe siendo negativo, lo que evidencia que todavía no se capturan plenamente las relaciones más complejas presentes en los datos.

El gráfico de comparación temporal (Figura 37) resalta cómo el modelo Ridge, especialmente tras la optimización del hiperparámetro α , logra seguir mejor las tendencias generales del precio del carbono en comparación con los enfoques previos. Sin embargo, se observa que el modelo aún presenta limitaciones para capturar las fluctuaciones más finas y complejas de los datos, lo que refuerza la necesidad de explorar alternativas más avanzadas.



Figura 37: Comparación temporal entre los valores reales y las predicciones del modelo Ridge optimizado

Los residuos del modelo Ridge optimizado, representados en la Figura 38, muestran una distribución menos dispersa y con valores más próximos a cero en comparación con los modelos analizados previamente. Esto sugiere que el modelo optimizado logra capturar mejor las relaciones entre las variables predictoras y la variable objetivo, aunque aún persisten ciertos errores sistemáticos en las predicciones.

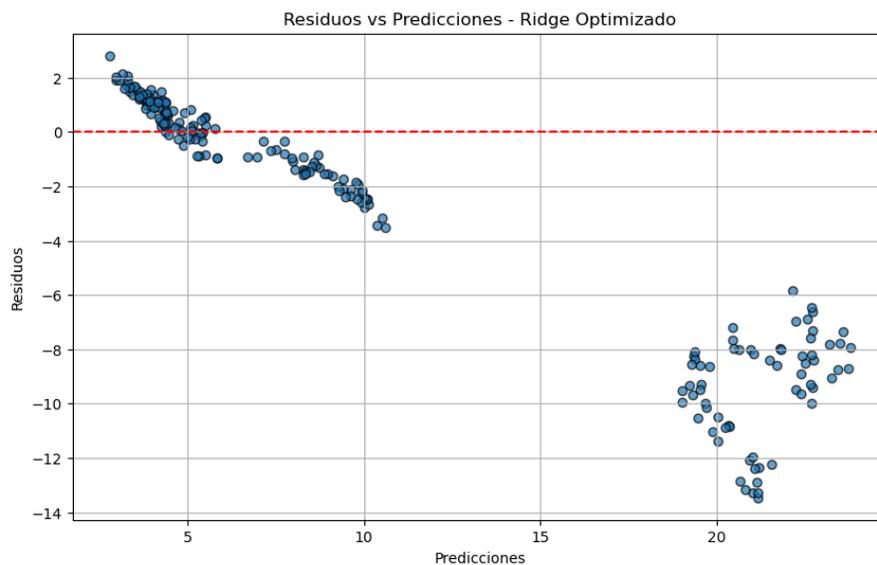


Figura 38: Relación entre los residuos y las predicciones del modelo Ridge optimizado

La Figura 39, que ilustra el histograma de la distribución de los residuos, evidencia una disminución en los valores extremos en comparación con el modelo Ridge sin optimizar. Aunque la distribución de los residuos no es completamente simétrica ni normal, el comportamiento observado refleja una mejora notable en el ajuste del modelo, con una mayor concentración de residuos alrededor de cero.

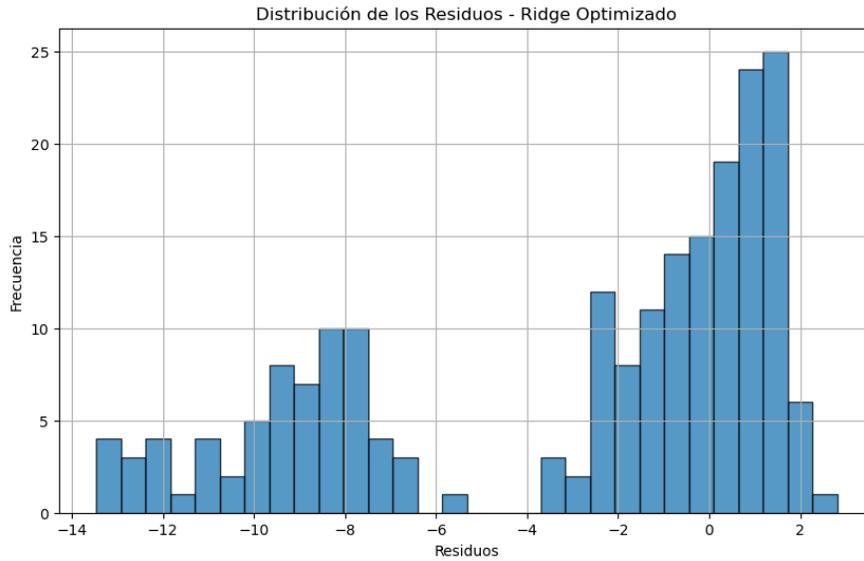


Figura 39: Histograma de la distribución de los residuos del modelo Ridge optimizado

En la Figura 40, que compara los valores reales y predichos, se observa que las predicciones del modelo optimizado, si bien no se alinean perfectamente con los valores reales, presentan una menor dispersión en comparación con los modelos anteriores. Esto indica que el modelo Ridge optimizado ofrece un ajuste más preciso y razonable, acercándose más a la línea ideal $y = x$, lo que refuerza su desempeño superior tras la optimización del hiperparámetro α .

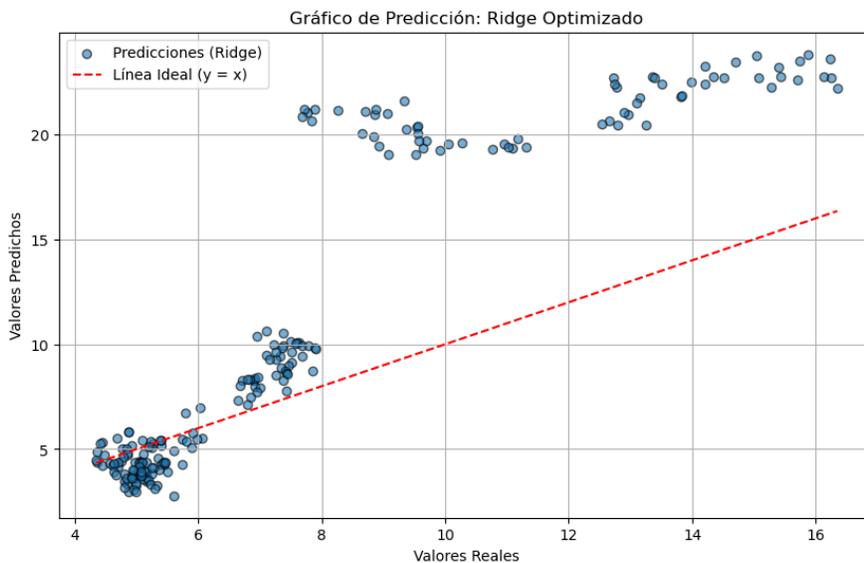


Figura 40: Comparación entre los valores reales y las predicciones del modelo Ridge optimizado

El modelo Ridge, tras su optimización, ha demostrado ser más eficaz que su versión inicial y otros enfoques lineales evaluados previamente. La reducción en el RMSE (de 6.1887 a 5.5331) y la mejora en el R^2 indican que la regularización es un elemento clave para abordar problemas de sobreajuste y multicolinealidad. Sin embargo, el desempeño aún limitado sugiere que la estructura lineal del modelo no es suficiente para capturar la complejidad de los datos y que podrían ser necesarias alternativas no lineales para obtener predicciones más precisas.

6.4. Árbol de Decisión

Un árbol de decisión es un modelo de aprendizaje supervisado utilizado tanto para clasificación como para regresión. Este enfoque se basa en la partición recursiva de los datos en subconjuntos más pequeños, organizados en una estructura jerárquica de nodos y ramas. Los árboles de decisión son ampliamente utilizados debido a su simplicidad, capacidad interpretativa y su habilidad para manejar datos tanto categóricos como continuos.

La estructura de un árbol de decisión es la siguiente:

- **Nodo Raíz:** Representa la totalidad del conjunto de datos y la característica inicial utilizada para realizar la primera partición.
- **Nodos Internos:** Representan particiones sucesivas del conjunto de datos en función de los valores de las variables predictoras.
- **Hojas:** Corresponden a los resultados finales, que son las predicciones realizadas por el modelo.

La construcción de un árbol de decisión se basa en seleccionar iterativamente las variables predictoras que mejor separen los datos. Para ello, se utiliza una métrica de división, como las siguientes:

Para clasificación:

- **Gini Impurity:** Evalúa qué tan bien se separan las clases en un nodo.
- **Entropía:** Mide la pureza de los datos dentro de un nodo.

Para regresión:

- **Reducción del Error Cuadrático (MSE):** Busca minimizar la varianza dentro de los nodos.

El algoritmo sigue los siguientes pasos:

1. Elegir la mejor característica (variable predictora) para dividir los datos en función de la métrica seleccionada.
2. Dividir los datos en dos o más subconjuntos basados en la característica seleccionada.
3. Repetir el proceso para cada subconjunto hasta que se cumpla un criterio de parada, como:
 - Alcanzar un número mínimo de muestras en los nodos.
 - Profundidad máxima del árbol.
 - La ganancia de información adicional es insignificante.

En regresión, la calidad de una división se evalúa mediante la minimización de la suma de los errores cuadrados:

$$MSE_{división} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2$$

Donde:

- y_i es el valor real de la variable dependiente.
- \hat{y} es el valor promedio de la variable dependiente en el nodo.

El algoritmo selecciona la división que minimice el MSE.

Para evitar el sobreajuste, los árboles de decisión incluyen hiperparámetros que regulan su complejidad:

- **Profundidad Máxima (max_depth):** Limita el número de divisiones jerárquicas permitidas.
- **Mínimo Número de Muestras en un Nodo (min_samples_split):** Especifica el número mínimo de datos necesarios para dividir un nodo.
- **Mínimo Número de Muestras en una Hoja (min_samples_leaf):** Determina el tamaño mínimo permitido para los nodos hoja.

- **Criterio (criterion):** Para regresión, suele ser “squared_error” o “mae”.

Para un nodo k, el valor predicho por el árbol de decisión es la media de los valores en ese nodo que sigues la siguiente fórmula:

$$\hat{y}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} y_i$$

Donde:

- n_k es el número de muestras en el nodo
- y_i son los valores reales de la variable dependiente en el nodo.

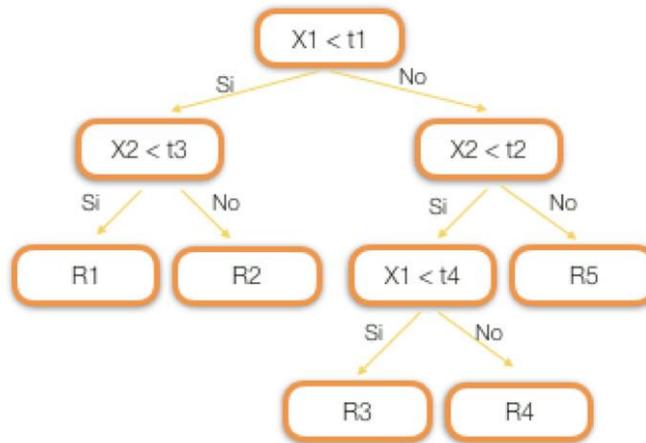


Figura 41: Estructura de un Árbol de Decisión [50]

En el siguiente análisis se implementará un modelo de árbol de decisión para predecir el precio del carbono en el mercado EU ETS. Se analizarán los resultados obtenidos, evaluando su precisión y capacidad de generalización, además de optimizar sus hiperparámetros para obtener el mejor rendimiento posible. En este análisis, el modelo de árbol de decisión fue evaluado utilizando dos configuraciones: una inicial sin optimización y otra optimizada para maximizar su rendimiento predictivo.

```
Evaluación del modelo sin optimizar:  
Root Mean Squared Error (RMSE): 16.5795  
R2 Score: -23.4914
```

Output 13: Métricas modelo Árbol de Decisión

La configuración inicial del modelo sin optimización arrojó un RMSE de 16.5795 y un R² Score de -23.4914. Aunque el modelo pudo capturar algunas relaciones entre las variables predictoras y el precio del carbono, su desempeño general es limitado, como lo indica el R² negativo, que refleja una incapacidad para superar un modelo trivial basado en el promedio de los datos.

Para mejorar el rendimiento del modelo, se llevó a cabo una búsqueda de hiperparámetros utilizando una combinación de valores para:

- La profundidad máxima del árbol (`max_depth`).
- El número mínimo de muestras requeridas para dividir un nodo (`min_samples_split`).
- El número mínimo de muestras que debe contener un nodo hoja (`min_samples_leaf`).

El proceso de optimización identificó los siguientes hiperparámetros óptimos:

- **`max_depth`: 10**
- **`min_samples_split`: 10**
- **`min_samples_leaf`: 1**

Con esta configuración, el modelo optimizado logró un RMSE de 16.1807 y un R² Score de -22.3272 en el conjunto de prueba. Si bien las métricas aún indican una limitación para capturar toda la variabilidad de los datos, el modelo optimizado muestra una ligera mejora en comparación con la versión inicial.

```
Evaluación del mejor modelo en el conjunto de prueba:  
Root Mean Squared Error (RMSE): 16.1807  
R2 Score: -22.3272
```

Output 14: Métricas modelo Árbol de Decisión optimizado

En la Figura 43 muestra el árbol de decisión optimizado, donde se aprecia cómo las divisiones jerárquicas en las variables predictoras contribuyen a estimar el precio del carbono. Este gráfico

destaca que variables como el precio del gas, el tipo de cambio USD/EUR y el precio del petróleo son fundamentales para la predicción. La estructura del árbol ilustra la importancia de estas variables y las interacciones clave entre ellas.

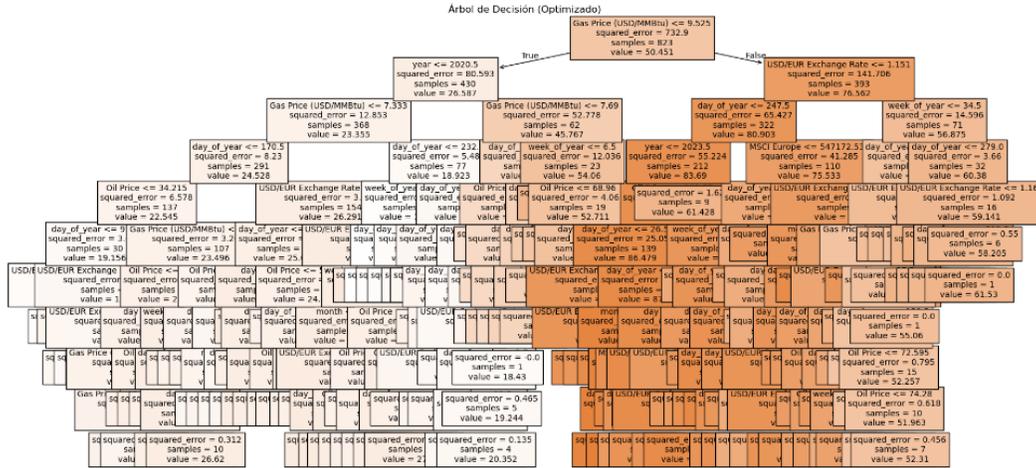


Figura 42: Visualización del Árbol de Decisión optimizado

En la Figura 44, se comparan los residuos frente a las predicciones para ambos modelos (sin optimizar y optimizado). El modelo optimizado muestra una menor dispersión en los residuos, lo que sugiere una mayor estabilidad y capacidad para capturar patrones subyacentes en los datos.

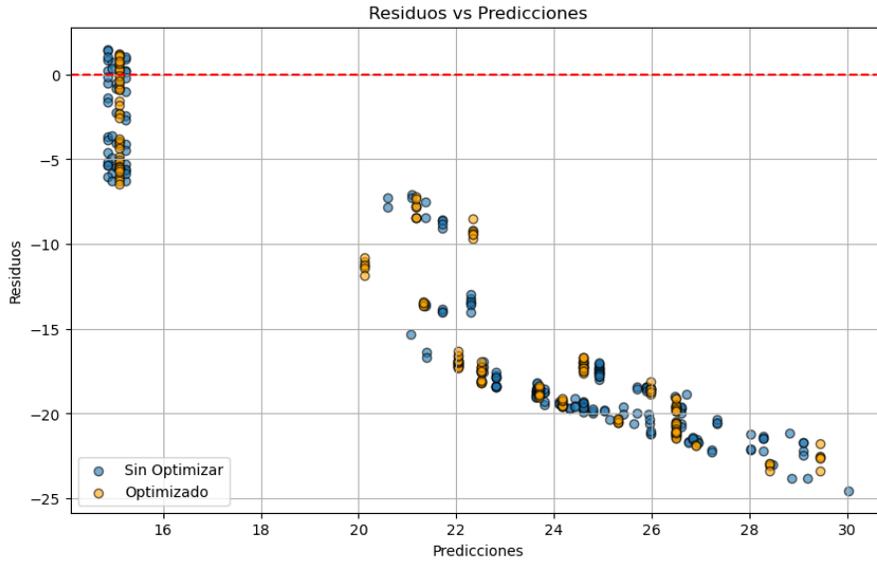


Figura 43: Relación entre los residuos y las predicciones de los modelos Árbol de Decisión

En la Figura 45, se presentan los residuos comparativos entre el modelo inicial y el optimizado. La reducción de la dispersión de los residuos en el modelo optimizado indica una mejora en el ajuste general. Aunque todavía hay valores significativos alejados de cero, la mayor concentración de residuos cercanos a la línea base refuerza la efectividad de la optimización.

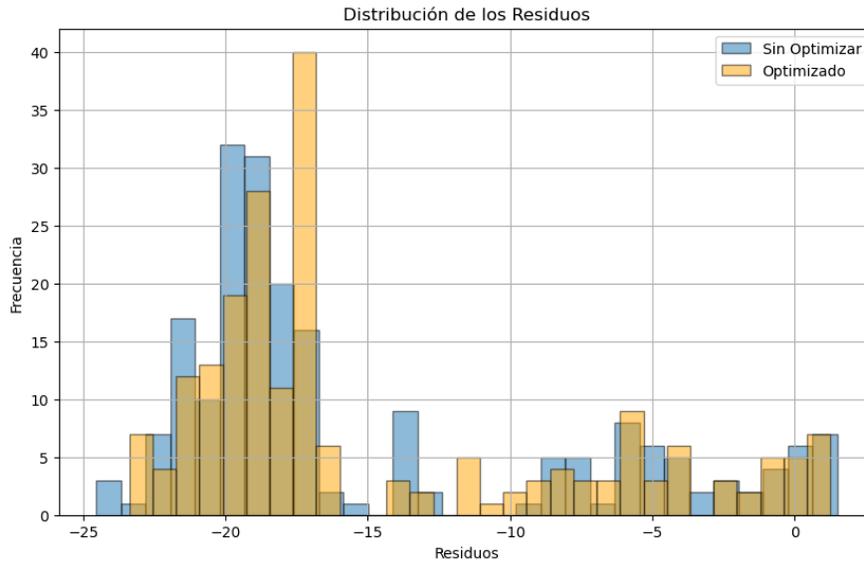


Figura 44: Histograma de la distribución de los residuos de los modelos Árbol de Decisión

Finalmente, en la Figura 46 muestra la relación entre los valores reales y las predicciones del modelo optimizado. Si bien los puntos predichos no se alinean perfectamente con la línea ideal ($y = x$), el modelo optimizado muestra una menor dispersión en comparación con el modelo sin optimizar, lo que refleja una mejora en la capacidad predictiva.

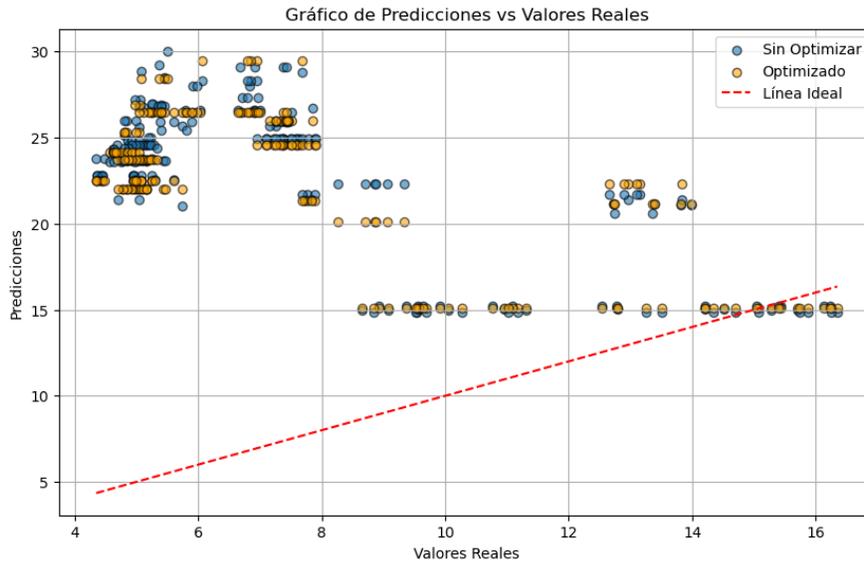


Figura 45: Comparación entre los valores reales y las predicciones de los modelos Árbol de Decisión

La Figura 46 compara temporalmente los valores reales y las predicciones realizadas por el modelo de árbol de decisión optimizado. El gráfico destaca una alineación más cercana entre las predicciones y los valores reales en comparación con la configuración inicial, evidenciando una mejora significativa tras la optimización. Sin embargo, aún se observan discrepancias notables en ciertos tramos, lo que refleja que, aunque el modelo optimizado capta mejor las relaciones subyacentes en los datos, sigue siendo insuficiente para capturar toda la complejidad inherente al problema.

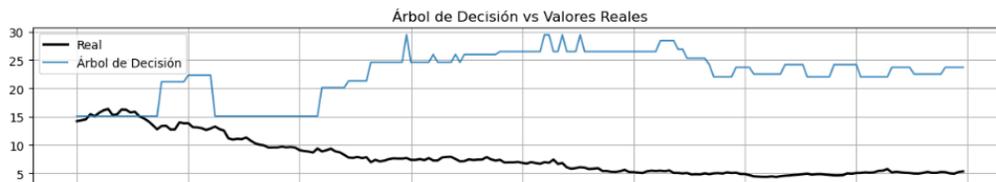


Figura 46: Comparación temporal entre los valores reales y las predicciones del modelo de Árbol de Decisión optimizado

El modelo de árbol de decisión optimizado mejora su desempeño respecto a la configuración inicial, como lo evidencian las métricas de error y los análisis visuales. Sin embargo, el desempeño general del modelo sigue siendo limitado, con un RMSE de 16.1807 y un R^2 Score de -22.3272. Esto sugiere que, aunque el árbol de decisión puede capturar algunas relaciones significativas en los datos, no es suficiente para abordar completamente la complejidad del problema. En los siguientes apartados se explorarán modelos más avanzados, como ensambles y técnicas basadas en aprendizaje profundo, para evaluar si pueden superar estas limitaciones y proporcionar predicciones más precisas.

6.5. Random Forest

El modelo de Random Forest es un algoritmo basado en árboles de decisión que utiliza un enfoque de aprendizaje por ensamble. Este método combina múltiples árboles de decisión independientes para mejorar la precisión y la estabilidad de las predicciones. El principio fundamental detrás de Random Forest es que, al promediar los resultados de varios árboles, se pueden mitigar los problemas de sobreajuste y obtener un modelo más robusto.

El algoritmo de Random Forest construye múltiples árboles de decisión utilizando diferentes subconjuntos del conjunto de datos, seleccionados mediante un proceso de muestreo aleatorio con reemplazo, conocido como bagging. Cada árbol se entrena con una muestra distinta, y dentro de cada uno, las divisiones se realizan utilizando un subconjunto aleatorio de características. Este enfoque introduce variabilidad entre los árboles y reduce la correlación entre ellos, lo que mejora la robustez del modelo.

En cuanto a la predicción, para problemas de regresión, el modelo calcula el promedio de las predicciones realizadas por todos los árboles, lo que proporciona un resultado final más estable. Por otro lado, para problemas de clasificación, el modelo selecciona la clase más frecuente entre las predicciones de los árboles, consolidando una decisión mayoritaria.

Una característica destacada del Random Forest es su capacidad para evaluar la importancia de las variables predictoras. Esto se logra al medir cuánto contribuye cada característica a la mejora de la pureza de los nodos, como la reducción del error, durante el proceso de entrenamiento. Este

análisis permite identificar cuáles variables tienen un impacto más significativo en las predicciones del modelo.

Para problemas de regresión, la predicción final del modelo es el promedio de las predicciones de los árboles individuales que siguen la siguiente ecuación:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N \hat{y}_i$$

Donde:

- \hat{y} es la predicción final.
- N es el número total de árboles en el modelo.
- \hat{y}_i es la predicción del árbol i .

Para problemas de clasificación, la predicción se basa en la clase más frecuente entre todos los árboles:

$$y = \text{moda}(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N)$$

El desempeño del modelo puede mejorarse ajustando los hiperparámetros clave:

- **Número de árboles (`n_estimators`):** Determina cuántos árboles se construirán. Un número mayor generalmente mejora la precisión, aunque aumenta el costo computacional.
- **Profundidad máxima de los árboles (`max_depth`):** Limita la complejidad de los árboles individuales para evitar sobreajuste.
- **Número de características seleccionadas (`max_features`):** Controla cuántas características se consideran para dividir cada nodo.
- **Mínimo número de muestras para dividir (`min_samples_split`):** Establece el número mínimo de muestras necesarias para dividir un nodo.
- **Mínimo número de muestras en un nodo hoja (`min_samples_leaf`):** Determina el tamaño mínimo de los nodos hoja.

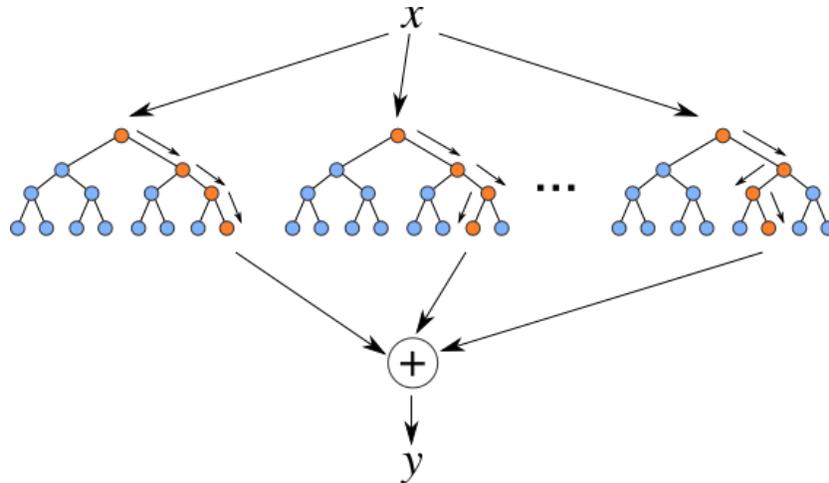


Figura 47: Estructura del modelo Random Forest [51]

Con esta base teórica, Random Forest se presenta como una extensión más poderosa y robusta del modelo de árbol de decisión, capaz de manejar relaciones complejas y reducir la varianza en las predicciones. Esto lo hace particularmente adecuado para datos complejos como los analizados en este trabajo.

El modelo Random Forest fue evaluado inicialmente con su configuración base, obteniendo un RMSE de 14.7101 y un R^2 de -18.2798. Estos valores muestran un desempeño mejorado respecto a otros modelos previamente analizados, pero el coeficiente de determinación negativo indica que el modelo base aún no logra capturar adecuadamente la complejidad de los datos, superando apenas el rendimiento de un modelo trivial que predijera el promedio.

```
Evaluación del modelo base de Random Forest:
Root Mean Squared Error (RMSE): 14.7101
R2 Score: -18.2798
```

Output 15: Métricas modelo Random Forest

Para mejorar el rendimiento del modelo, se llevó a cabo un proceso de optimización de hiperparámetros. En este ajuste se consideraron:

- **Número de estimadores (número de árboles):** Se fijó en 300, lo que asegura una mayor estabilidad en las predicciones al promediar un mayor número de resultados individuales.

- **Profundidad máxima:** Sin restricciones, permitiendo que los árboles se expandan lo suficiente para capturar la mayor cantidad de patrones posibles en los datos.
- **Muestras mínimas para dividir un nodo:** Se mantuvo en 2, permitiendo la división de nodos incluso con pocas observaciones.
- **Muestras mínimas en hojas:** Fijada en 1, asegurando que los nodos terminales puedan contener al menos una muestra, maximizando la granularidad de las predicciones.

Con estos hiperparámetros optimizados, el modelo Random Forest alcanzó un RMSE de 14.4227 y un R^2 de -17.5337 en el conjunto de prueba. La reducción del RMSE indica que el modelo optimizado logró capturar patrones en los datos con mayor precisión, minimizando los errores promedio en las predicciones. Si bien el coeficiente R^2 sigue siendo negativo, muestra una ligera mejora respecto al modelo base, destacando el impacto positivo de la optimización.

```
Evaluación del modelo optimizado de Random Forest:  
Root Mean Squared Error (RMSE): 14.4227  
R2 Score: -17.5337
```

Output 16: Métricas modelo Random Forest optimizado

En la Figura 49, se presenta un ejemplo de un árbol individual dentro del bosque de Random Forest optimizado. Este árbol ilustra cómo se generan las divisiones en función de las características más relevantes y cómo las decisiones se toman de manera jerárquica. Aunque Random Forest opera promediando las predicciones de múltiples árboles, analizar uno de ellos permite comprender la lógica subyacente del modelo. Las variables como el precio del gas (USD/MMBtu), la tasa de cambio USD/EUR y el precio del petróleo emergen como características clave en las divisiones principales.

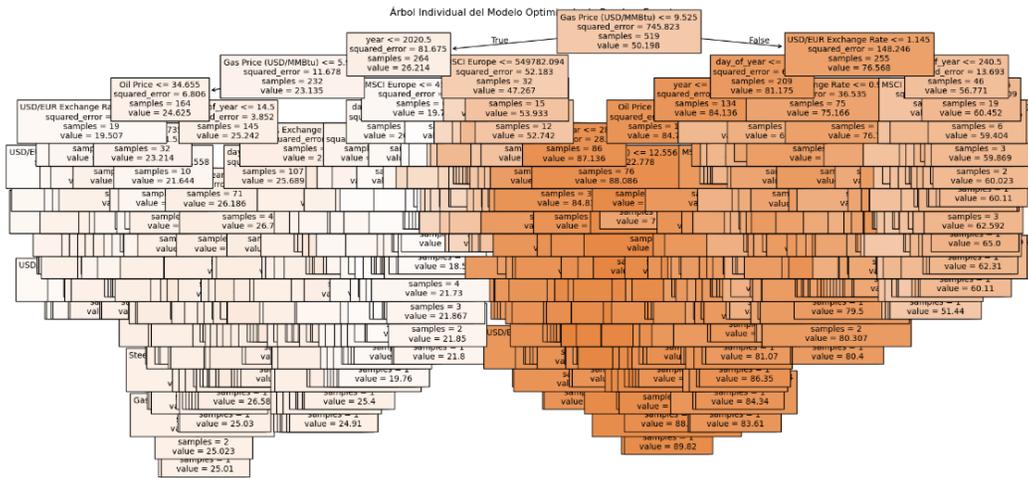


Figura 48: Visualización de un árbol individual del modelo Random Forest optimizado

La Figura 50 muestra la relación entre los residuos y las predicciones generadas por los modelos base y optimizado. En ambos casos, los residuos tienden a alejarse de cero en los valores extremos de las predicciones, pero el modelo optimizado muestra una menor dispersión en torno a la línea horizontal (residuo = 0). Esto sugiere que las predicciones optimizadas están mejor ajustadas a los valores reales, especialmente en los intervalos medios.

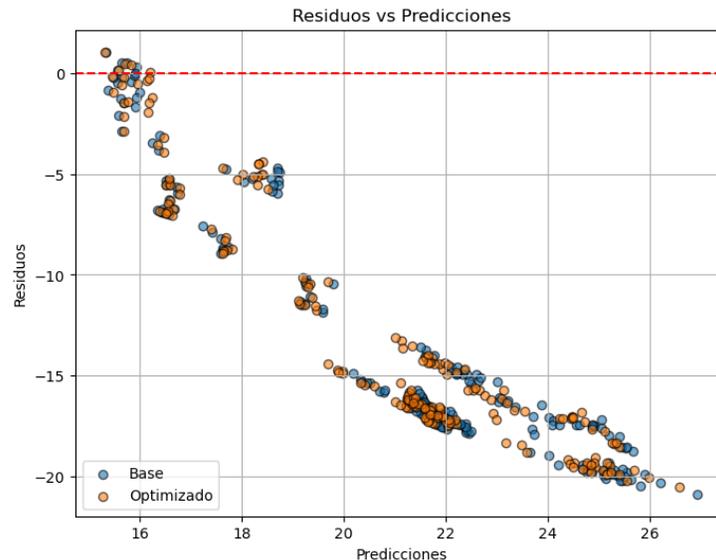


Figura 49: Relación entre los residuos y las predicciones de los modelos Random Forest

La Figura 51 compara la distribución de los residuos entre los modelos base y optimizado. Se observa que el modelo optimizado reduce la concentración de residuos en valores extremos negativos, trasladándolos hacia el centro, cerca de cero. Esto evidencia que el modelo optimizado es más efectivo para predecir valores dentro del rango de los datos observados. Sin embargo, todavía persisten residuos en regiones alejadas de cero, indicando que aún hay margen de mejora en el ajuste general.

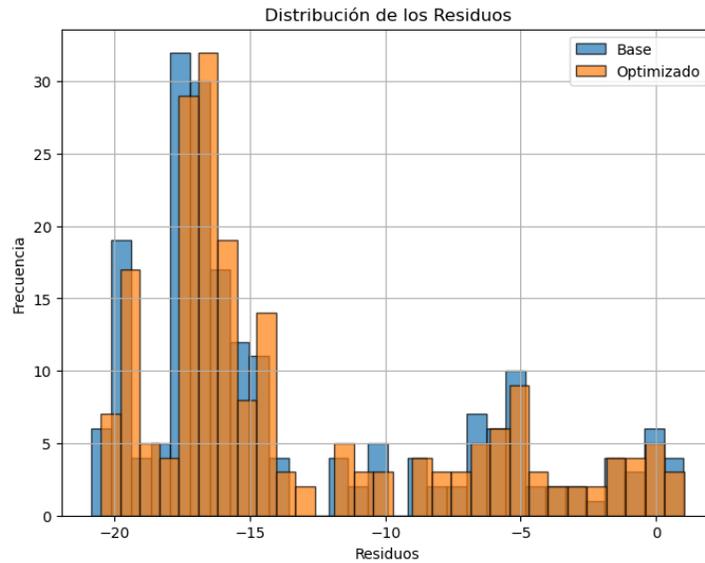


Figura 50: Histograma de la distribución de los residuos de los modelos Random Forest

En la Figura 52, se analiza la alineación entre los valores reales y las predicciones para ambos modelos. Aunque ninguna configuración logra una correspondencia perfecta con la línea ideal (donde los valores predichos son iguales a los reales), el modelo optimizado muestra una menor dispersión en comparación con el modelo base. Esto refuerza la idea de que la optimización contribuyó a un ajuste más preciso de los datos.

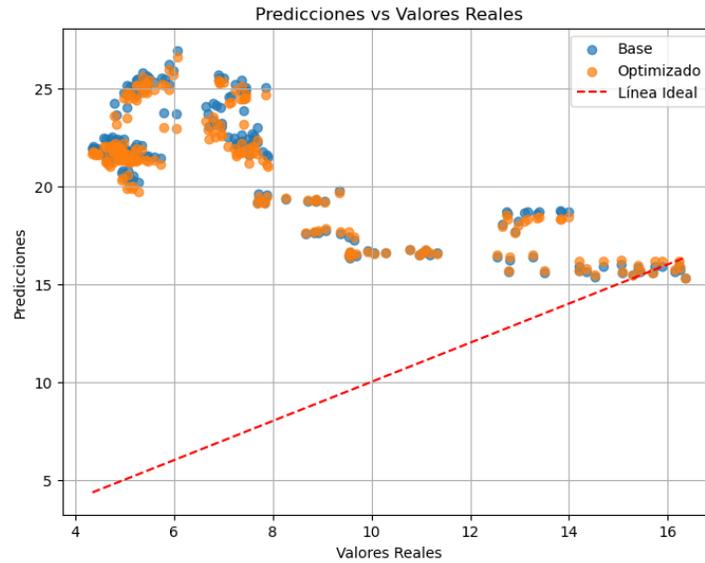


Figura 51: Comparación entre los valores reales y las predicciones de los modelos Random Forest

La Figura 52 presenta una comparación temporal entre los valores reales y las predicciones del modelo Random Forest optimizado. Aunque se observa una mayor alineación en tramos específicos, el modelo sigue fallando en capturar adecuadamente las fluctuaciones más abruptas del precio del carbono. Este comportamiento pone de manifiesto que, aunque la optimización mejoró el ajuste general, el modelo aún carece de la capacidad para representar completamente la dinámica del conjunto de datos.

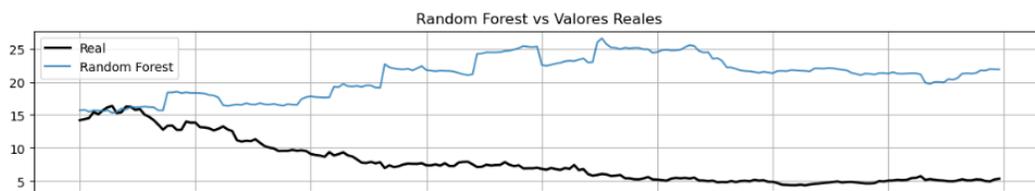


Figura 52: Comparación temporal entre los valores reales y las predicciones del modelo Random Forest optimizado

El modelo Random Forest muestra ciertas ventajas frente al árbol de decisión simple, principalmente en términos de reducción del error y una mayor capacidad para manejar datos complejos. Tras la optimización, el RMSE de Random Forest (14.4227) es ligeramente inferior al del árbol de decisión optimizado (16.1807), lo que indica un ajuste más preciso de los datos. Sin embargo, ambos modelos comparten un coeficiente de determinación (R^2) negativo, lo que

demuestra que, aunque Random Forest logra capturar algo más de la variabilidad de los datos, su utilidad práctica sigue siendo limitada en este contexto.

En comparación con modelos como Ridge y Lasso, Random Forest maneja mejor las relaciones no lineales, pero la diferencia en rendimiento no es suficientemente significativa como para compensar su mayor complejidad. Aunque Random Forest introduce el concepto de bagging, que reduce la varianza del modelo al combinar múltiples árboles, el R^2 negativo (-17.5337) refleja que incluso este enfoque no es capaz de superar las limitaciones de los datos o el problema de predicción planteado.

Frente al árbol de decisión, Random Forest aporta una mejora clave: la estabilidad y reducción del sobreajuste inherente a un solo árbol. Esto es evidente en los gráficos de residuos, donde los valores extremos son menos frecuentes en el modelo Random Forest. Sin embargo, la brecha entre ambos modelos sigue siendo moderada, y el rendimiento general de ambos es insuficiente para ofrecer predicciones confiables.

Un aspecto positivo de Random Forest es su capacidad para calcular la importancia de las características, lo que proporciona insights valiosos sobre los factores más influyentes en el precio del carbono, como los precios del gas y el petróleo. Este análisis interpretativo puede ser útil desde una perspectiva exploratoria, pero no logra compensar las limitaciones del modelo en términos de precisión predictiva.

6.6. Gradient Boosting

El algoritmo Gradient Boosting se basa en la construcción secuencial de modelos débiles, típicamente árboles de decisión de baja profundidad, que se combinan de forma aditiva para formar un modelo fuerte. El proceso comienza con una inicialización básica en la que se utiliza un modelo simple como punto de partida, como por ejemplo la predicción del valor medio de la variable objetivo en el conjunto de entrenamiento. A partir de este modelo inicial, se calculan los residuos, que representan la diferencia entre las predicciones actuales y los valores reales. Estos residuos identifican las áreas donde el modelo actual tiene un rendimiento deficiente y requieren corrección.

En cada iteración, un nuevo árbol de decisión se ajusta para predecir los residuos calculados previamente. Este árbol tiene la función de capturar las relaciones no explicadas por el modelo actual. Una vez ajustado, el modelo general se actualiza al agregar las predicciones del nuevo árbol a las predicciones existentes. Este ajuste se realiza utilizando un factor de aprendizaje (η) que controla la contribución de cada modelo débil al modelo final. Este enfoque permite que el modelo avance gradualmente en la dirección que reduce el error de manera controlada. Este proceso de ajuste de residuos y actualización del modelo se repite durante un número definido de iteraciones o hasta que el error deje de mejorar significativamente, según lo estipulado por el criterio de parada.

La ecuación general del modelo puede expresarse como:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x)$$

Donde:

- **$F_m(x)$** : es el modelo en la iteración m .
- **$F_{m-1}(x)$** : es el modelo en la iteración anterior.
- **η** : es el factor de aprendizaje.
- **$h_m(x)$** : es el modelo débil (árbol de decisión) ajustado a los residuos en la iteración m .

Gradient Boosting es propenso al sobreajuste si no se regula adecuadamente. Por ello, incluye varios mecanismos de regularización que mejoran su robustez y capacidad de generalización:

- **Factor de aprendizaje (η)**: Controla la tasa a la que el modelo se actualiza en cada iteración. Valores más pequeños reducen la probabilidad de sobreajuste, pero requieren más iteraciones para converger.
- **Número de iteraciones ($n_estimators$)**: Define cuántos árboles se entrenarán. Un número excesivo puede causar sobreajuste, mientras que un número bajo puede llevar a un modelo subajustado.
- **Profundidad de los árboles (max_depth)**: Limita la complejidad de cada árbol para evitar que los modelos débiles se conviertan en modelos fuertes, lo que podría comprometer la idea de ensamble.
- **Tasa de subsampling**: Controla la proporción de muestras y características utilizadas en cada iteración, lo que introduce diversidad en el ensamble y mejora la generalización.

Gradient Boosting difiere de Random Forest en que los árboles no se construyen de manera independiente, sino de forma secuencial, corrigiendo los errores de los árboles anteriores. Esto permite que Gradient Boosting se enfoque en áreas específicas del espacio de datos, logrando un ajuste más preciso en comparación con el promedio generalizado de Random Forest. Sin embargo, esta secuencialidad también lo hace más propenso al sobreajuste y más lento en términos de entrenamiento.

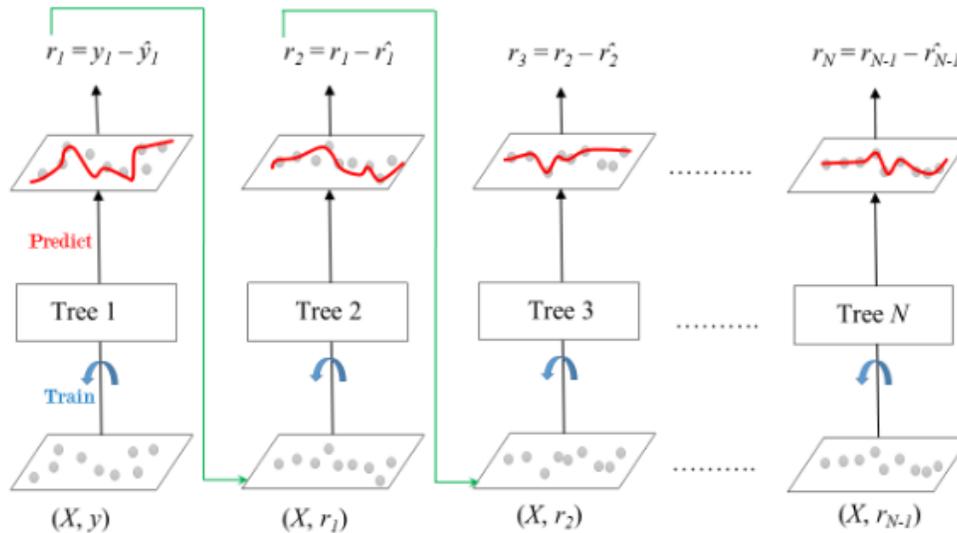


Figura 53: Proceso secuencial del algoritmo Gradient Boosting [52]

El modelo base de Gradient Boosting (GBM) alcanzó un RMSE de 12.9967, lo que supone una mejora considerable respecto a los modelos anteriormente evaluados, como el Árbol de Decisión y Random Forest. Esta mejora en el error de predicción indica que GBM tiene una mayor capacidad para capturar patrones en los datos. No obstante, el coeficiente de determinación (R^2) fue de -14.0500, lo que sigue siendo negativo, evidenciando que el modelo aún no logra explicar adecuadamente la variabilidad de los datos. Esto subraya que, aunque el GBM es más robusto que los enfoques anteriores, los desafíos inherentes a este conjunto de datos limitan su efectividad en su configuración base.

Modelo Base GBM:
Root Mean Squared Error (RMSE): 12.9967
R² Score: -14.0500

Output 17: Métricas modelo Gradient Boosting

El modelo base de Gradient Boosting (GBM) alcanzó un RMSE de 12.9967, lo que supone una mejora considerable respecto a los modelos anteriormente evaluados, como el Árbol de Decisión y Random Forest. Esta mejora en el error de predicción indica que GBM tiene una mayor capacidad para capturar patrones en los datos. No obstante, el coeficiente de determinación (R²) fue de -14.0500, lo que sigue siendo negativo, evidenciando que el modelo aún no logra explicar adecuadamente la variabilidad de los datos. Esto subraya que, aunque el GBM es más robusto que los enfoques anteriores, los desafíos inherentes a este conjunto de datos limitan su efectividad en su configuración base.

Para mejorar el rendimiento del modelo GBM, se realizó una optimización manual de los hiperparámetros. Este proceso consistió en explorar diversas combinaciones de los principales parámetros que controlan el funcionamiento del modelo. En particular, se ajustaron:

- **Número de estimadores:** Este parámetro define la cantidad de árboles individuales que componen el modelo. Tras evaluar diferentes configuraciones, se encontró que 300 estimadores proporcionaban el mejor balance entre precisión y generalización.
- **Tasa de aprendizaje:** Este factor controla la contribución de cada árbol al modelo final. Una tasa de aprendizaje de 0.1 demostró ser efectiva para capturar patrones en los datos sin sobreajustarse.
- **Profundidad máxima:** La profundidad de los árboles afecta directamente la capacidad del modelo para capturar relaciones complejas. En este caso, una profundidad máxima de 3 fue suficiente para encontrar un equilibrio entre complejidad y generalización.
- **Submuestreo:** Este parámetro determina la fracción de muestras utilizadas para construir cada árbol, lo que introduce aleatoriedad y reduce el sobreajuste. Un valor de 0.8 proporcionó buenos resultados en este caso.

Tras implementar estos ajustes, el modelo optimizado logró un RMSE de 12.1561, lo que representa una mejora significativa respecto al modelo base. Además, el R² aumentó ligeramente,

alcanzando -12.1663. Aunque el R^2 sigue siendo negativo, lo que indica que el modelo aún no es plenamente efectivo, estas métricas reflejan que la optimización permitió capturar mejor la complejidad de los datos.

```
Modelo Optimizado GBM:  
Root Mean Squared Error (RMSE): 12.1561  
R2 Score: -12.1663
```

Output 18: Métricas modelo Gradient Boosting optimizado

La Figura 55 presenta una comparación temporal entre los valores reales y las predicciones generadas por el modelo Gradient Boosting optimizado. A pesar de las mejoras obtenidas, las predicciones del modelo no logran capturar completamente las fluctuaciones observadas en los valores reales, especialmente en los extremos del rango. Este comportamiento pone en evidencia las limitaciones del modelo, que aunque mejora sustancialmente el ajuste general respecto a versiones previas, sigue sin ser completamente adecuado para representar la dinámica del precio del carbono.



Figura 54: Comparación temporal entre los valores reales y las predicciones del modelo Gradient Boosting optimizado

En la Figura 56, se presenta uno de los árboles individuales del modelo GBM optimizado. Este árbol muestra cómo las variables más importantes, como el precio del gas (USD/MMBtu) y la tasa de cambio USD/EUR, son seleccionadas en las divisiones iniciales. Estas variables juegan un papel clave en la reducción del error en el modelo, ya que sus valores reflejan relaciones significativas en el comportamiento del precio del carbono.

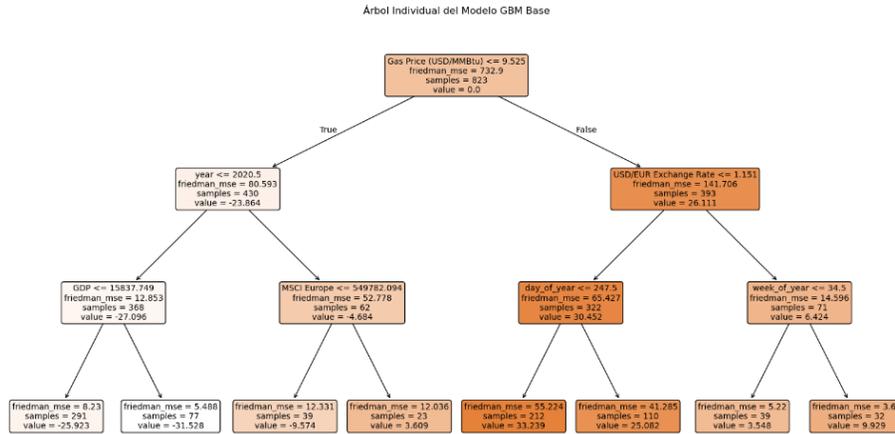


Figura 55: Visualización del árbol de decisión optimizado del modelo Gradient Boosting

En la Figura 57 se muestra la relación entre los residuos y las predicciones para ambos modelos. El modelo optimizado presenta una reducción en la magnitud de los residuos, particularmente en los extremos de las predicciones. Esta mejora sugiere que el modelo optimizado tiene una menor propensión al error en comparación con el modelo base, aunque persisten algunos residuos que no siguen un patrón claro.

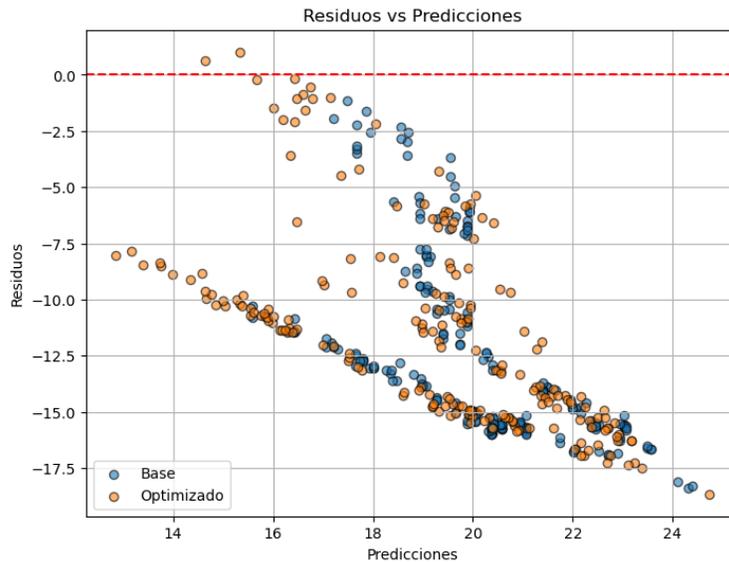


Figura 56: Relación entre los residuos y las predicciones de los modelos Gradient Boosting

La Figura 58 compara la distribución de los residuos entre el modelo base y el optimizado. En el modelo optimizado, los residuos están más concentrados alrededor de cero y muestran menos dispersión en comparación con el modelo base. Esto sugiere que el modelo optimizado tiene una mayor capacidad para predecir valores cercanos a los reales. Sin embargo, aún existen valores residuales extremos, lo que indica que el modelo no logra capturar ciertos patrones en los datos.

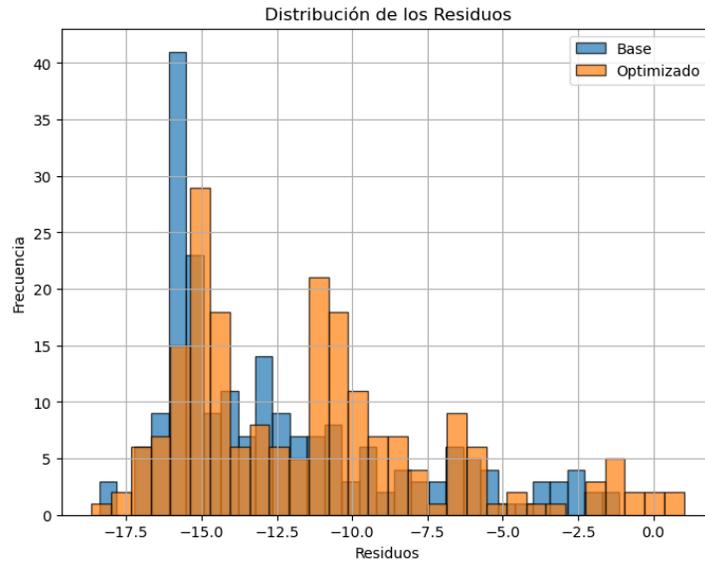


Figura 57: Distribución de los residuos de los modelos Gradient Boosting base y optimizado

La Figura 58 compara los valores reales con las predicciones de ambos modelos. El modelo optimizado muestra predicciones más cercanas a la línea ideal ($y = x$), especialmente en los valores intermedios. Sin embargo, en los extremos, tanto el modelo base como el optimizado presentan desviaciones significativas, lo que refleja las limitaciones del modelo para capturar la variabilidad en estos puntos.

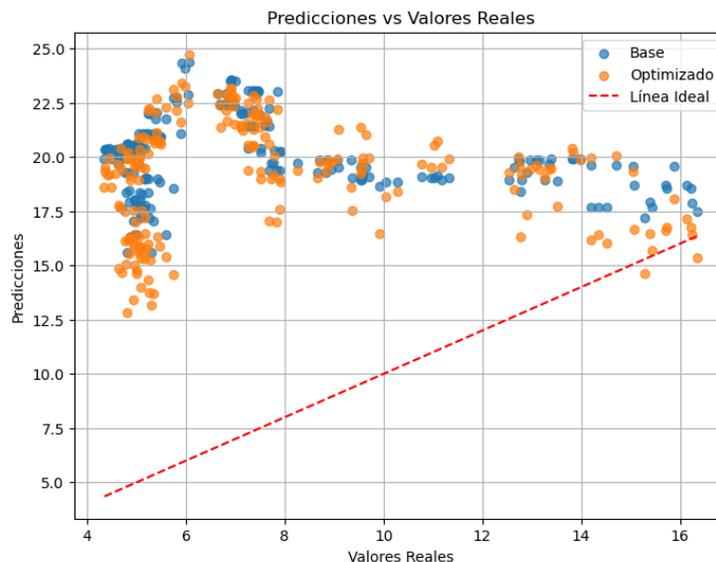


Figura 58: Comparación entre los valores reales y las predicciones de los modelos Gradient Boosting

En comparación con el Árbol de Decisión y Random Forest, el modelo Gradient Boosting ha demostrado un desempeño superior en términos de RMSE, alcanzando el error más bajo de todos los modelos evaluados hasta ahora. Esto evidencia la capacidad de Gradient Boosting para manejar relaciones complejas y patrones no lineales en los datos. Sin embargo, al igual que en los modelos previos, el R^2 negativo indica que los datos presentan desafíos significativos para cualquier enfoque lineal o no lineal.

A diferencia del Árbol de Decisión, que sufre de una alta varianza en su versión base y optimizada, y Random Forest, que aunque mejora la estabilidad no logra explicar bien los datos, Gradient Boosting combina de manera efectiva múltiples predicciones para lograr un modelo más robusto. Esto lo convierte en una mejora sustancial frente a los enfoques anteriores, aunque aún se encuentra lejos de ser útil para aplicaciones prácticas con este conjunto de datos.

El modelo Gradient Boosting, tanto en su versión base como optimizada, destaca por su capacidad para capturar patrones más complejos en los datos. Si bien logra el RMSE más bajo hasta ahora (12.1561 en su versión optimizada), el R^2 negativo (-12.1663) indica que todavía está lejos de explicar adecuadamente la variabilidad de los datos. Esto sugiere que, aunque Gradient Boosting representa un paso adelante en términos de precisión, los datos analizados continúan presentando desafíos significativos.

6.7. Extreme Gradient Boosting

XGBoost (Extreme Gradient Boosting) se basa en el concepto de boosting, una técnica de ensamble en la que los modelos se construyen de forma secuencial para corregir los errores cometidos por los modelos anteriores. En este enfoque, cada nuevo modelo busca reducir los errores residuales de los modelos previos, ajustándose a las discrepancias no resueltas. El objetivo de XGBoost es minimizar una función de pérdida definida por el usuario, ajustando los parámetros del modelo para lograr predicciones más precisas.

Matemáticamente, el modelo puede ser representado como una suma de predictores, que generalmente son árboles de decisión. La función del modelo en el paso t se define como:

$$\hat{y}_i = \sum_{k=1}^t f_k(x_i), \quad f_k \in \mathcal{F},$$

Donde \mathcal{F} representa el espacio de todos los posibles árboles de decisión. El objetivo del modelo es encontrar los árboles f_k que minimicen la función de pérdida regularizada L :

$$L = \sum_{i=1}^n \ell(y_i, \hat{y}_i) + \sum_{k=1}^t \Omega(f_k)$$

Donde:

- $\ell(y_i, \hat{y}_i)$ es la función de pérdida, como el error cuadrático medio (MSE) para problemas de regresión o la entropía cruzada para problemas de clasificación.
- $\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2$ es un término de regularización que penaliza la complejidad del modelo. Aquí, T es el número de hojas del árbol, λ es un parámetro de regularización, y w son los pesos de las hojas.

Durante el entrenamiento, XGBoost utiliza una aproximación de segundo orden para optimizar la función de pérdida. En lugar de trabajar directamente con la función de pérdida, el modelo utiliza una expansión de Taylor de segundo orden para aproximarla, lo que mejora la estabilidad y la convergencia del modelo. La actualización de los parámetros se basa en los gradientes g_i (primera derivada) y los hessianos h_i (segunda derivada) de la función de pérdida:

$$L^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

XGBoost realiza la construcción de árboles de forma eficiente mediante una técnica llamada “exact greedy algorithm”, que busca dividir los datos en los puntos que minimicen la pérdida. Para optimizar el proceso, el modelo utiliza procesamiento paralelo y gestión eficiente de memoria. Además, XGBoost soporta manejo nativo de datos faltantes, asignando estos valores a las ramas de los árboles donde generan la menor pérdida.

Un aspecto destacado de XGBoost es su capacidad para incluir regularización tanto L1 (Lasso) como L2 (Ridge), lo que ayuda a evitar el sobreajuste al limitar la complejidad de los árboles y mejorar la capacidad de generalización del modelo. Asimismo, permite trabajar con diferentes funciones de pérdida, como el error absoluto, el logarítmico o funciones personalizadas, adaptándose a las necesidades del problema.

Por último, XGBoost también proporciona métricas para evaluar la importancia de las características utilizadas en el modelo. Esto se logra al medir cuánto contribuye cada característica a reducir la función de pérdida durante la construcción de los árboles. Esta información es útil no solo para interpretar el modelo, sino también para identificar las variables más relevantes en el conjunto de datos.

En este apartado, se presenta el análisis de desempeño del modelo XGBoost aplicado al conjunto de datos del precio del carbono. A diferencia de los modelos previos, como Gradient Boosting, XGBoost utiliza optimizaciones avanzadas en la construcción y ajuste de los árboles de decisión, lo que lo posiciona como una de las herramientas más robustas para el aprendizaje supervisado. En particular, se destacan características como la regularización L1 y L2, el manejo eficiente de memoria, y la construcción de árboles mediante técnicas de optimización basadas en gradientes de segundo orden.

El modelo fue inicialmente entrenado con sus parámetros predeterminados, obteniendo una línea base de desempeño para ser utilizada como punto de comparación. Posteriormente, se realizó una búsqueda manual de hiperparámetros para optimizar el modelo y mejorar su capacidad predictiva. Los ajustes realizados incluyen parámetros como el número de árboles (`n_estimators`), la profundidad de los árboles (`max_depth`), la tasa de aprendizaje (`learning_rate`), la fracción de

muestras utilizadas por cada árbol (subsample), y la fracción de características seleccionadas por árbol (colsample_bytree). Además, se añadieron términos de regularización para reducir el riesgo de sobreajuste, considerando tanto L1 (reg_alpha) como L2 (reg_lambda).

El modelo base de XGBoost, entrenado sin optimización de parámetros, logró un Root Mean Squared Error (RMSE) de 15.5713 y un R² Score de -20.6032. Estas métricas reflejan que el modelo base es capaz de captar ciertos patrones generales, pero aún no logra ajustarse adecuadamente a las relaciones complejas de los datos. La alta magnitud del RMSE y el valor negativo de R² evidencian que el modelo base tiene un desempeño inferior al de modelos previamente analizados, como Gradient Boosting, y requiere ajustes para mejorar su capacidad predictiva.

```
Modelo Base XGBoost:  
Root Mean Squared Error (RMSE): 15.5713  
R2 Score: -20.6032
```

Output 19: Métricas modelo Extreme Gradient Boosting

La optimización de XGBoost se centró en ajustar los hiperparámetros mediante una búsqueda manual, evaluando cada configuración en el conjunto de prueba. Esto permitió encontrar la combinación de parámetros que minimiza el error de predicción.

Los hiperparámetros optimizados incluyen:

- **Número de estimadores (n_estimators):** Se incrementó a 500, permitiendo al modelo construir más árboles y corregir errores con mayor precisión.
- **Profundidad máxima (max_depth):** Establecida en 3, lo que garantiza un balance entre la captura de relaciones complejas y la generalización del modelo.
- **Tasa de aprendizaje (learning_rate):** Ajustada a 0.2, lo que permitió que el modelo aprendiera patrones más rápidamente, reduciendo los errores en iteraciones tempranas.
- **Fracción de muestras (subsample) y fracción de características (colsample_bytree):** Ambas definidas en 0.8, lo que introduce variabilidad en cada árbol y reduce la correlación entre ellos.
- **Regularización (reg_alpha y reg_lambda):** Se utilizó un valor de reg_lambda=5 para regularizar los coeficientes de los árboles y limitar su complejidad.

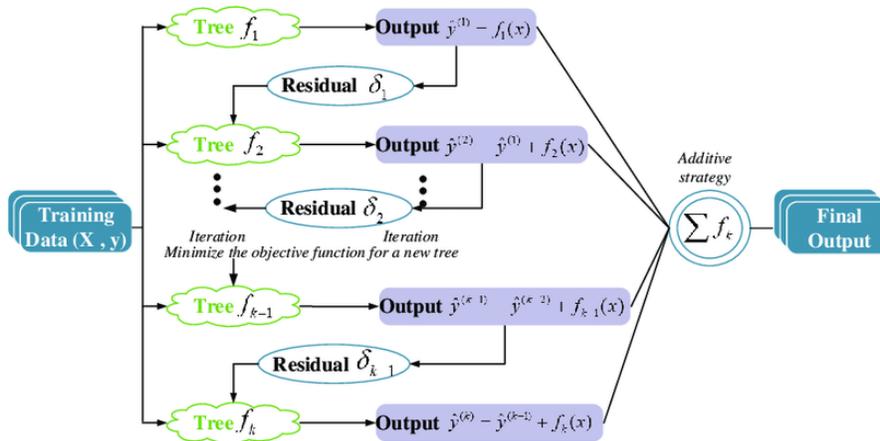


Figura 59: Estructura del algoritmo XGBoost: construcción secuencial y optimización [53]

Con estos ajustes, el modelo optimizado alcanzó un RMSE de 11.6919 y un R² Score de -11.1799, logrando superar significativamente el desempeño del modelo base.

Modelo Optimizado XGBoost:
Root Mean Squared Error (RMSE): 11.6919
R² Score: -11.1799

Output 20: Métricas modelo Extreme Gradient Boosting optimizado

La Figura 61 presenta una comparación temporal entre los valores reales y las predicciones generadas por el modelo XGBoost optimizado. A pesar de la mejora en las métricas de error, las predicciones del modelo no logran ajustarse completamente a los valores reales, especialmente en los extremos del rango. Esto evidencia las limitaciones del modelo para capturar ciertas relaciones complejas en los datos, aunque muestra un ajuste superior al de los modelos previamente analizados.



Figura 60: Comparación temporal entre los valores reales y las predicciones del modelo XGBoost optimizado

La gráfica muestra la relación entre los residuos y las predicciones para los modelos base y optimizado. En el modelo base, los residuos presentan una mayor dispersión respecto a la línea de referencia ($y = 0$), lo que refleja un ajuste menos preciso. En contraste, el modelo optimizado muestra residuos más concentrados cerca de la línea de referencia, evidenciando una mejora en la capacidad predictiva del modelo tras la optimización. Sin embargo, persisten patrones en los residuos que podrían indicar la presencia de relaciones no capturadas por el modelo.

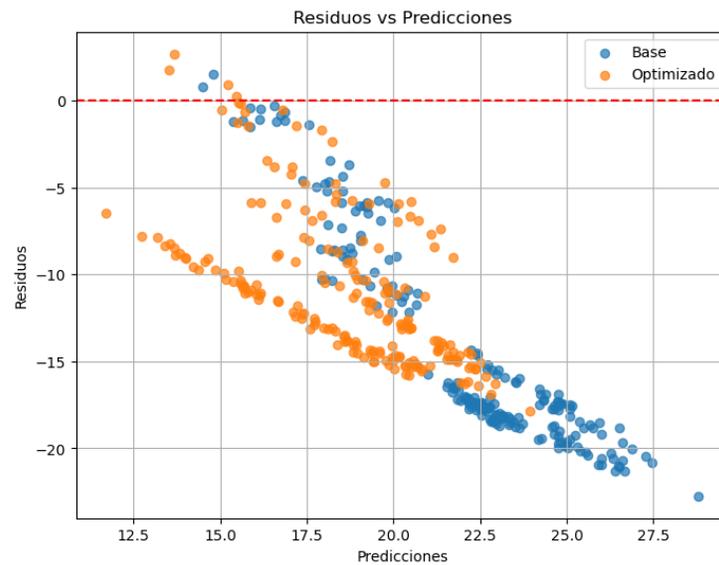


Figura 61: Relación entre los residuos y las predicciones de los modelos XGBoost

El histograma compara la distribución de los residuos entre el modelo base y el optimizado. En el modelo base, se observa una mayor frecuencia de residuos extremos, lo que indica un ajuste más deficiente. Por otro lado, el modelo optimizado logra una distribución más centrada alrededor de cero y con menos valores extremos, lo que sugiere una mejor generalización y una reducción del error absoluto promedio.

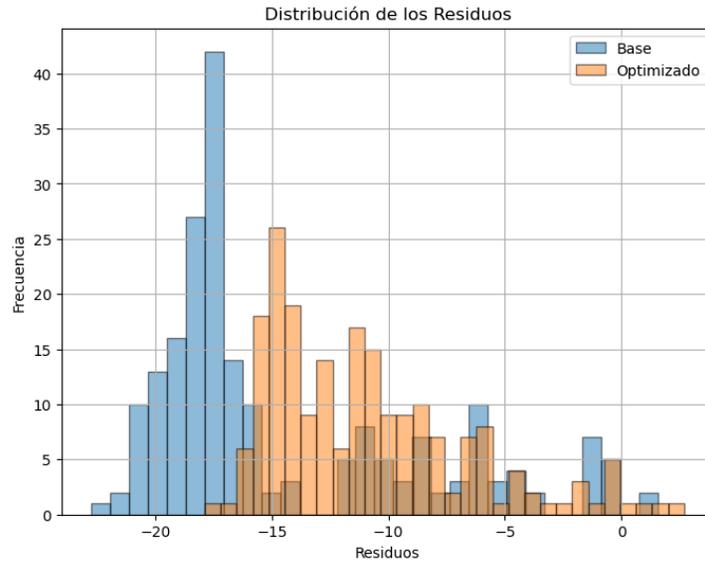


Figura 62: Histograma de la distribución de los residuos de los modelos XGBoost

Este gráfico ilustra la correspondencia entre los valores reales y las predicciones de los modelos base y optimizado. Mientras que el modelo base presenta una mayor dispersión alrededor de la línea ideal ($y = x$), el modelo optimizado muestra predicciones más cercanas a los valores reales, especialmente en el rango medio de los datos. Aunque el modelo optimizado logra una mejora notable, la desviación en los extremos del rango indica que sigue existiendo margen para reducir los errores en las predicciones de valores extremos.

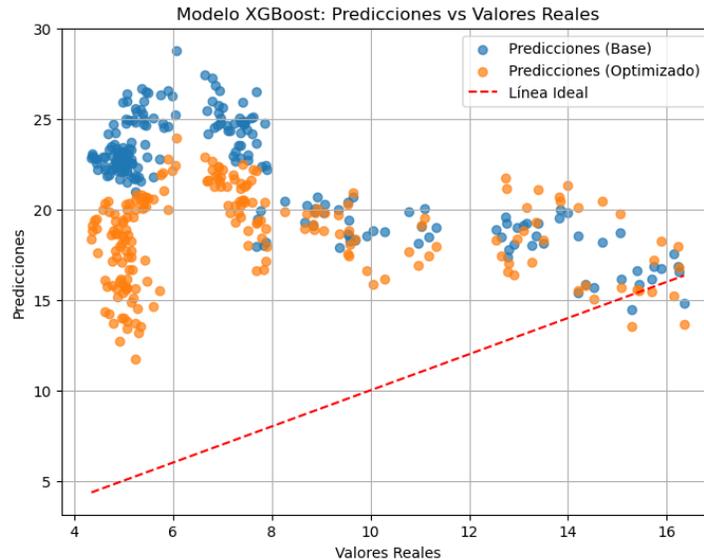


Figura 63: Comparación entre los valores reales y las predicciones de los modelos XGBoost

El modelo optimizado de XGBoost logró un RMSE inferior al de Gradient Boosting (11.6919 vs. 12.1561), demostrando su capacidad para capturar patrones más detallados en los datos. Este resultado refuerza el papel de XGBoost como una mejora iterativa de Gradient Boosting, beneficiándose de su regularización avanzada y técnicas de optimización.

No obstante, al igual que en modelos anteriores, el R^2 negativo indica que, aunque el error absoluto disminuye, la capacidad explicativa del modelo sigue siendo limitada. Esto sugiere que los datos presentan un nivel de complejidad que los modelos de boosting no logran resolver completamente, destacando la necesidad de explorar métodos más sofisticados o enfoques híbridos.

El modelo XGBoost optimizado representa un paso adelante en el análisis de los precios del carbono, destacando por su capacidad para manejar relaciones complejas y reducir significativamente el error de predicción. Aunque sigue siendo limitado por un negativo, el RMSE más bajo alcanzado hasta ahora indica que este modelo logra un mejor ajuste en comparación con los abordajes anteriores. Sin embargo, la complejidad de los datos exige considerar modelos adicionales o técnicas avanzadas en futuros análisis para abordar las limitaciones actuales.

6.8. Multilayer Perceptron (MLP)

En esta sección, se introduce el Multilayer Perceptron (MLP) como una técnica avanzada de aprendizaje supervisado basada en redes neuronales artificiales. Dado que este es el primer modelo basado en redes neuronales de este trabajo, es importante comenzar explicando los conceptos fundamentales de estas arquitecturas y cómo se aplican al problema de predicción del precio del carbono.

Las redes neuronales artificiales (ANN, por sus siglas en inglés) son algoritmos inspirados en el cerebro humano, diseñados para modelar relaciones complejas entre entradas y salidas. Estas estructuras consisten en un conjunto de nodos (neuronas) organizados en capas, donde cada nodo está conectado a otros mediante pesos, que son ajustados durante el proceso de aprendizaje. El objetivo principal de una ANN es aprender patrones subyacentes en los datos, incluso cuando las relaciones entre las variables predictoras y la variable objetivo no son lineales ni evidentes. Esto las convierte en herramientas poderosas para tareas como clasificación, predicción y detección de anomalías.

El Multilayer Perceptron (MLP) es uno de los modelos más básicos y fundamentales dentro de las redes neuronales artificiales. Se caracteriza por tener al menos tres capas:

1. **Capa de Entrada:** Recibe las variables predictoras (features) como entrada.
2. **Capas Ocultas:** Realizan transformaciones no lineales mediante funciones de activación, permitiendo capturar patrones complejos en los datos.
3. **Capa de Salida:** Produce la predicción final, que en este caso es un valor continuo debido a que estamos resolviendo un problema de regresión.

Matemáticamente, el funcionamiento del MLP puede representarse de la siguiente manera:

$$y = f\left(\sum_{i=1}^n w_i \cdot x_i + b\right)$$

Donde:

- x_i son las características de entrada.
- w_i son los pesos asociados a cada entrada.

- **b** es el término de sesgo (bias), que permite al modelo ajustar las predicciones cuando las entradas son cero.
- **f** es la función de activación aplicada al resultado de la suma ponderada.

En una red neuronal multicapa, esta operación se repite para cada nodo en todas las capas ocultas, propagando la información hacia adelante hasta la capa de salida. En cada capa, la salida de las neuronas se convierte en la entrada de la siguiente capa. Esta propagación permite al modelo aprender representaciones jerárquicas de los datos, capturando relaciones complejas que no son evidentes en las características originales.

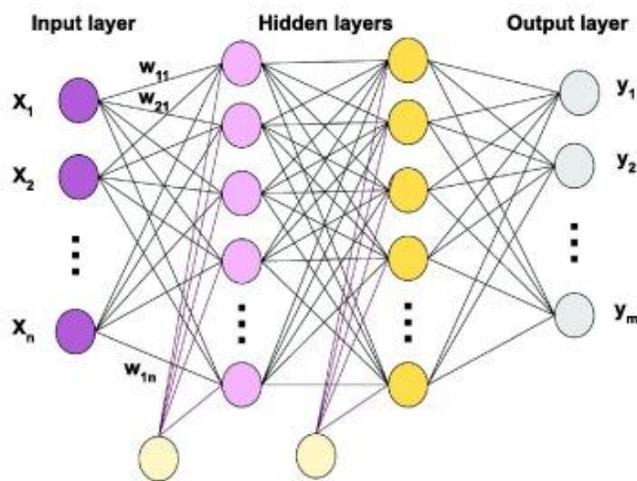


Figura 64: Estructura de una Red Neuronal Multicapa (Multilayer Perceptron) [54]

La estructura del MLP lo convierte en un modelo versátil, adecuado para problemas como el análisis de series temporales o la predicción de variables dependientes altamente no lineales. En este caso, se emplea para predecir el precio del carbono, un problema caracterizado por relaciones no triviales y patrones difíciles de identificar con enfoques más tradicionales.

En este apartado, se analiza el rendimiento de los modelos Multilayer Perceptron (MLP) utilizados para la predicción del precio del carbono. Se desarrollaron dos enfoques principales: un modelo base no optimizado y un modelo optimizado tras realizar un ajuste exhaustivo de los hiperparámetros. Este análisis incluye una descripción detallada de los modelos, una discusión sobre los resultados obtenidos y la presentación de figuras ilustrativas para facilitar la comprensión del rendimiento de cada modelo.

El modelo base MLP fue diseñado como una arquitectura básica, con la finalidad de establecer un punto de referencia inicial. La configuración del modelo incluyó:

- **Una primera capa oculta** con 64 neuronas y la función de activación ReLU, que introduce no linealidad al modelo y le permite capturar patrones complejos en los datos.
- **Una segunda capa oculta** con 32 neuronas, también utilizando ReLU, para agregar mayor capacidad de representación.
- **Una capa de salida** con activación lineal, adecuada para problemas de regresión, donde se busca predecir valores continuos.

Para mitigar el sobreajuste, se incluyó un Dropout del 20% en las capas ocultas, eliminando aleatoriamente algunas conexiones durante el entrenamiento. El modelo utilizó el optimizador Adam con una tasa de aprendizaje fija de 0.001 y se entrenó durante 100 épocas, con una validación cruzada del 20% y callbacks como EarlyStopping y ReduceLROnPlateau para gestionar el aprendizaje y prevenir un entrenamiento excesivo.

Los resultados del modelo no han sido los esperados, puesto que el modelo mostró limitaciones significativas al intentar capturar la complejidad del problema. Los valores obtenidos para las métricas clave fueron:

- **Error cuadrático medio de la raíz (RMSE):** 15.19.
- **Coefficiente de determinación (R^2):** -19.54.

Estos resultados indican que el modelo no logró identificar adecuadamente los patrones subyacentes en los datos y que sus predicciones estaban lejos de los valores reales.

La Figura 65, presentada a continuación, muestra las predicciones del modelo frente a los valores reales, evidenciando una discrepancia significativa.

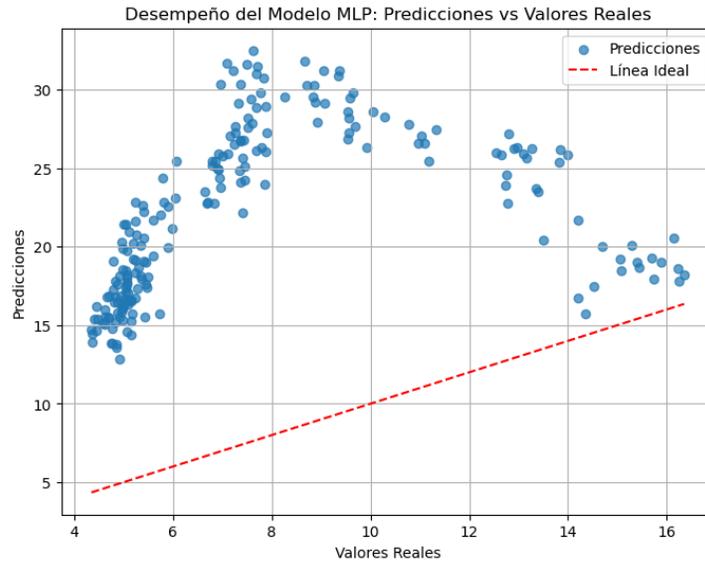


Figura 65: Desempeño del modelo MLP no optimizado: predicciones vs valores reales

Esta figura ilustra cómo las predicciones del modelo no optimizado presentan una alta dispersión con respecto a la línea ideal (línea roja), que representa una predicción perfecta.

Por otro lado, la Figura 66 muestra la evolución de la pérdida durante el entrenamiento y la validación del modelo. En este gráfico se observa que, aunque el modelo converge, los valores de pérdida permanecen altos, lo que refleja un ajuste inadecuado a los datos.

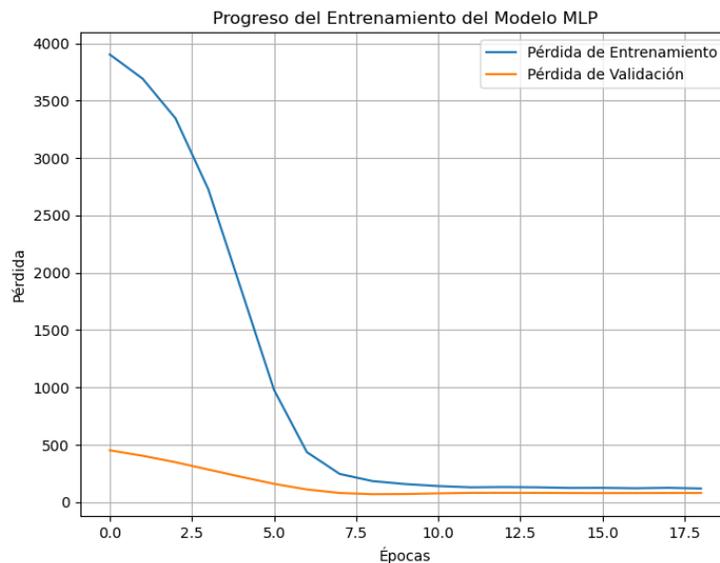


Figura 66: Progreso del Entrenamiento del modelo MLP no optimizado

En esta figura, se observa que tanto la pérdida de entrenamiento como la de validación disminuyen inicialmente, pero se estabilizan en valores elevados, lo que sugiere que el modelo no es capaz de reducir significativamente el error.

Dado el rendimiento limitado del modelo no optimizado, se implementó un proceso de optimización de hiperparámetros utilizando Keras Tuner, que permite explorar automáticamente combinaciones de configuraciones para encontrar aquellas que maximizan el rendimiento del modelo. Entre los parámetros ajustados, se incluyen:

- **El número de neuronas por capa:** Se probaron valores entre 64 y 128 para la primera capa y entre 32 y 64 para la segunda.
- **Las tasas de Dropout:** Se evaluaron valores entre 10% y 20% para reducir el sobreajuste.
- **La tasa de aprendizaje:** Se exploraron valores entre 0.001 y 0.01.

El modelo final optimizado incluyó:

- **Una primera capa oculta** con 64 neuronas.
- **Una segunda capa oculta** con 32 neuronas.
- **Hiperparámetros ajustados** que incluían un Dropout de 10% y una tasa de aprendizaje cercana a 0.009.

Este modelo fue entrenado durante 20 épocas, reduciendo el tiempo de entrenamiento en comparación con el modelo base.

Los resultados del modelo optimizado mostraron mejoras significativas en comparación con el modelo no optimizado:

- **RMSE:** 6.80.
- **R²:** -3.11.

Si bien el modelo optimizado logró reducir el error de las predicciones, los resultados aún indican que las relaciones subyacentes entre las variables predictoras y el precio del carbono no fueron capturadas completamente.

La Figura 67 muestra cómo las predicciones del modelo optimizado se acercan más a los valores reales en comparación con el modelo no optimizado.

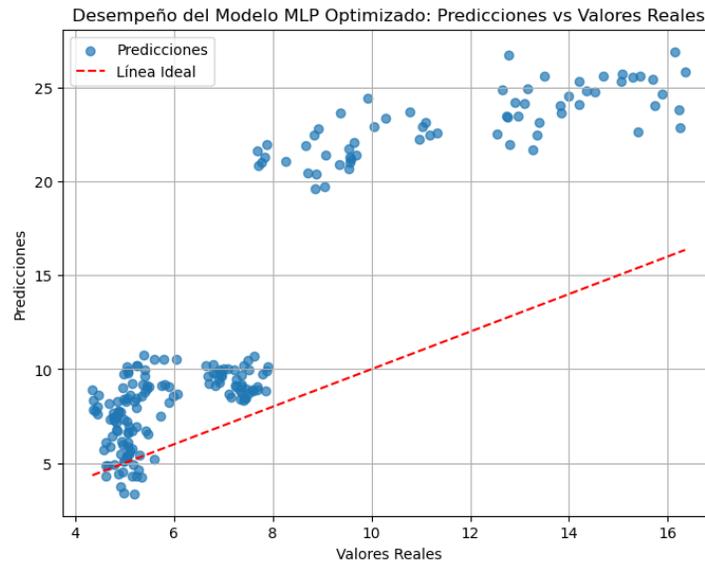


Figura 67: Desempeño del modelo MLP optimizado: predicciones vs valores reales

Este gráfico evidencia que las predicciones generadas por el modelo optimizado están más alineadas con los valores reales, aunque persisten errores significativos.

La Figura 68, por su parte, presenta la evolución de la pérdida durante el entrenamiento del modelo optimizado. Este gráfico destaca una convergencia más rápida y a valores de pérdida más bajos en comparación con el modelo no optimizado.

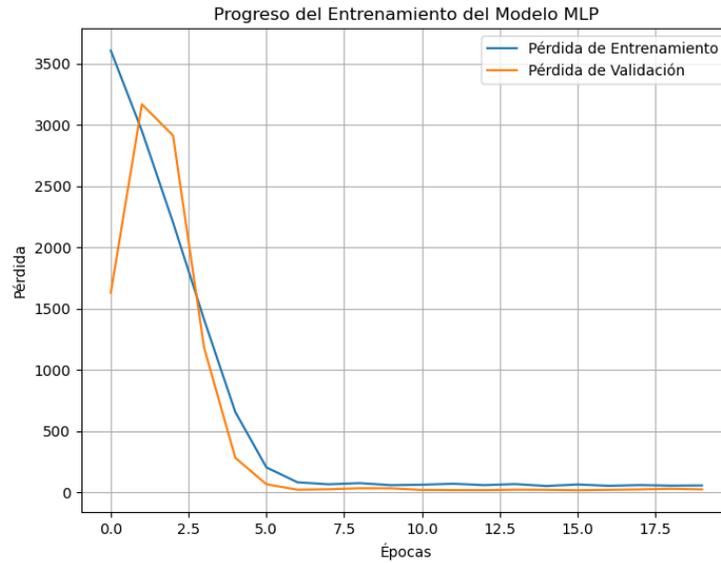


Figura 68: Progreso del Entrenamiento del modelo MLP optimizado

Esta figura muestra que el ajuste de los hiperparámetros permitió al modelo optimizado aprender más eficientemente y alcanzar una menor pérdida, tanto en el conjunto de entrenamiento como en el de validación.

Para analizar mejor las diferencias entre ambos enfoques, la Figura 5 presenta una comparación directa entre las predicciones de los modelos (no optimizado y optimizado) frente a los valores reales. En esta figura se observa cómo el modelo optimizado sigue más de cerca la tendencia de los datos reales, aunque aún existen discrepancias importantes.

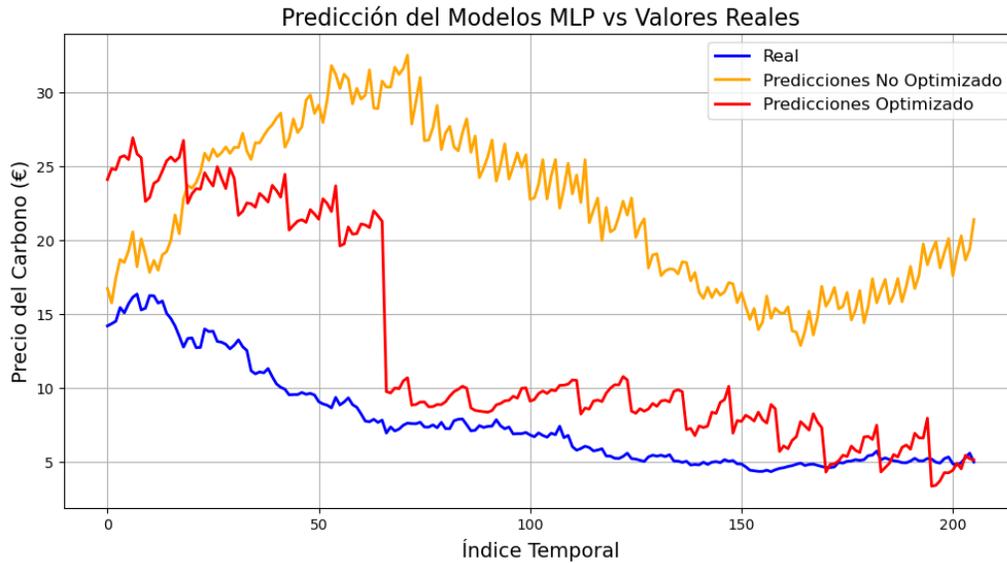


Figura 69: Predicción de los modelos MLP vs valores reales

En esta figura, se incluye la línea azul que representa los valores reales, junto con las predicciones del modelo no optimizado (línea naranja) y del modelo optimizado (línea roja). La comparación evidencia cómo la optimización mejora la precisión del modelo, pero también subraya la necesidad de explorar enfoques adicionales para capturar mejor la dinámica de los datos.

Este análisis destaca cómo el proceso de optimización de hiperparámetros puede mejorar significativamente el rendimiento de un modelo MLP. El modelo optimizado logró reducir el error y mejorar la precisión en comparación con el modelo base, aunque los resultados sugieren que pueden ser necesarias arquitecturas más avanzadas o datos adicionales para abordar la complejidad del problema de predicción del precio del carbono.

6.9. Long Short-Term Memory (LSTM)

Las redes neuronales recurrentes (RNN, por sus siglas en inglés) son un tipo especializado de redes neuronales diseñadas para manejar datos secuenciales, donde el orden y la relación temporal entre los datos son cruciales. A diferencia de las redes tradicionales que procesan datos de forma independiente, las RNN cuentan con conexiones recurrentes que permiten la propagación de información a través de pasos de tiempo. Esto las hace especialmente útiles para tareas como series

temporales, procesamiento de texto, audio y video. El principal problema de las RNN tradicionales es que tienden a sufrir de problemas como el desvanecimiento y explosión de gradientes, lo que limita su capacidad para capturar dependencias a largo plazo en los datos. Esto ocurre porque la información importante puede perderse o diluirse a medida que se propaga a través de muchas capas o pasos de tiempo.

Para superar estas limitaciones, Long Short-Term Memory (LSTM), una variante de las RNN, fue introducida por Hochreiter y Schmidhuber en 1997. Las LSTM están diseñadas específicamente para retener información relevante durante largos períodos de tiempo y desechar información innecesaria, gracias a su arquitectura única basada en “celdas de memoria” y “puertas de control”. Las LSTM son una poderosa herramienta en problemas donde las relaciones temporales o secuenciales son clave, como la predicción de series temporales, traducción automática y modelado de texto.

Cada unidad LSTM se basa en una celda de memoria que actúa como un “depósito” para almacenar información relevante a lo largo del tiempo. Además, se utilizan tres tipos de puertas para controlar el flujo de información hacia y desde la celda de memoria:

1. **Puerta de Olvido (Forget Gate):** Decide qué información debe descartarse de la celda de memoria. Esta puerta se calcula como:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Donde:

- **x_t :** entrada en el tiempo t
- **h_{t-1} :** salida del paso de tiempo anterior
- **W_f y b_f :** pesos y sesgos
- **σ :** función sigmoide, que escala los valores entre 0 y 1

2. **Puerta de Entrada (Input Gate):** Determina qué nueva información debe añadirse a la celda de memoria. Consiste en dos pasos:

- Calcular los valores candidatos para la nueva información:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

- Decidir cuánto de esta nueva información se almacena:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

3. **Puerta de Salida (Output Gate):** Decide qué información se debe extraer de la celda de memoria para producir la salida actual:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

- La salida final de la unidad se calcula como:

$$h_t = o_t \cdot \tanh(C_t)$$

Finalmente, la celda de memoria se actualiza con la información retenida y la nueva información seleccionada:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

Las redes LSTM presentan diversas ventajas que las convierten en una herramienta poderosa para el análisis de datos temporales y secuenciales. Una de sus principales fortalezas es su capacidad para capturar dependencias a largo plazo, ya que su arquitectura de puertas permite almacenar información relevante durante extensos períodos de tiempo sin sufrir problemas como el desvanecimiento de gradientes. Además, su flexibilidad las habilita para manejar secuencias de longitud variable y datos complejos, lo que las hace adecuadas para una amplia gama de aplicaciones. Entre estas destacan la predicción de series temporales, el procesamiento del lenguaje natural y el reconocimiento de voz, lo que evidencia su gran versatilidad.

Sin embargo, las LSTM también presentan ciertas desventajas que deben considerarse. Su arquitectura compleja implica un alto costo computacional, ya que el gran número de parámetros incrementa significativamente el tiempo de entrenamiento y los requisitos de hardware. Además, son difíciles de interpretar, pues funcionan como modelos de “caja negra”, lo que complica la comprensión de cómo toman decisiones específicas. Por último, su eficacia depende de la disponibilidad de grandes conjuntos de datos bien estructurados, lo que puede limitar su aplicación en contextos con datos insuficientes o mal preparados.

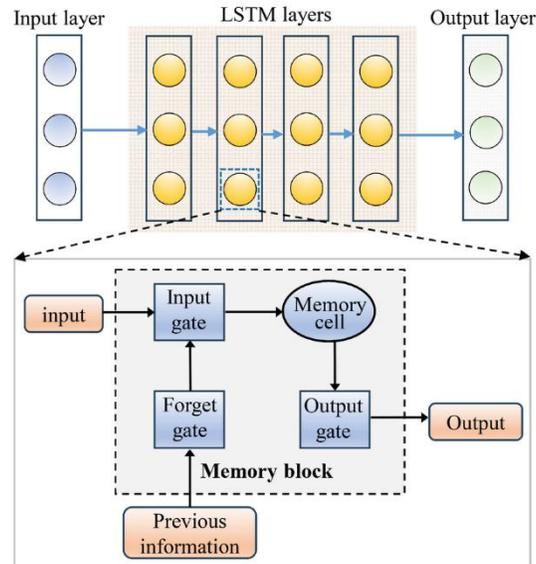


Figura 70: Estructura de una red LSTM [55]

En esta sección, se implementaron redes neuronales de tipo Long Short-Term Memory (LSTM) para abordar el problema de predicción de precios del carbono. Las LSTM son particularmente adecuadas para esta tarea debido a su capacidad para modelar relaciones temporales y secuenciales en los datos. Este tipo de red neuronal permite capturar patrones a largo plazo en series temporales, superando las limitaciones de las redes neuronales recurrentes tradicionales, como el desvanecimiento de gradientes. En este caso, las LSTM demostraron ser una herramienta poderosa para identificar las tendencias subyacentes en el mercado del carbono y realizar predicciones precisas basadas en datos históricos.

El modelo LSTM inicial fue diseñado con una arquitectura sencilla pero efectiva para evaluar su desempeño en la predicción de precios del carbono. Para comenzar, se realizó un escalado de los datos utilizando un MinMaxScaler para normalizarlos en el rango $[0,1]$, lo cual garantiza una mejor convergencia durante el entrenamiento. Se empleó una ventana deslizante de 60 días para generar las secuencias de entrada, lo que permitió al modelo aprender las dependencias temporales dentro de los datos históricos.

La arquitectura del modelo base incluyó dos capas LSTM: la primera configurada para retornar secuencias, lo que permitió que la segunda capa pudiera procesar información acumulada. Ambas capas estaban complementadas con una capa de Dropout para prevenir el sobreajuste, utilizando

una probabilidad de 0.2. Finalmente, una capa densa con una única neurona proporcionó las predicciones finales. El modelo fue compilado utilizando la función de pérdida de error cuadrático medio (MSE) y el optimizador Adam, conocido por su eficiencia en el ajuste de redes neuronales. El entrenamiento se llevó a cabo durante 100 épocas con un tamaño de lote de 32, y se emplearon técnicas de regularización como EarlyStopping y ReduceLROnPlateau para garantizar un entrenamiento efectivo.

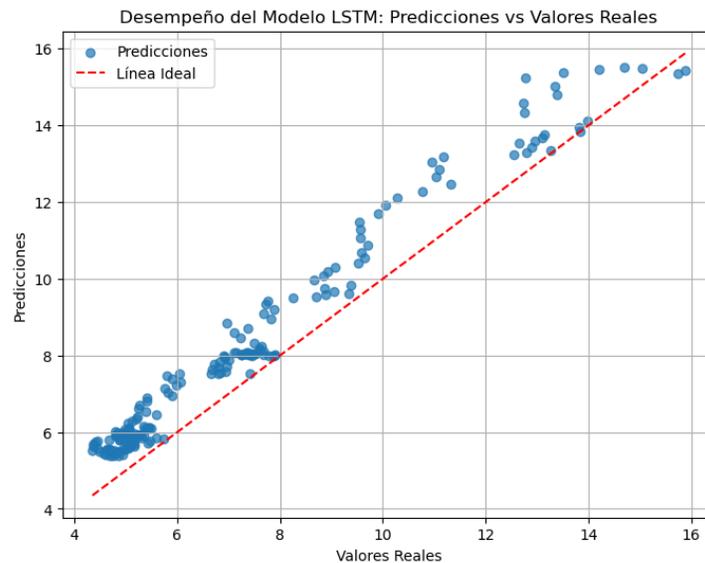


Figura 71: Desempeño del modelo LSTM: predicciones vs valores reales

El modelo base mostró resultados prometedores, como se observa en la Figura 71, que compara las predicciones del modelo con los valores reales. Aunque logró capturar tendencias generales, se identificaron algunas desviaciones que sugieren áreas de mejora. Los resultados cuantitativos del modelo base fueron los siguientes:

- **Root Mean Squared Error (RMSE):** 1.0262
- **R²:** 0.8686

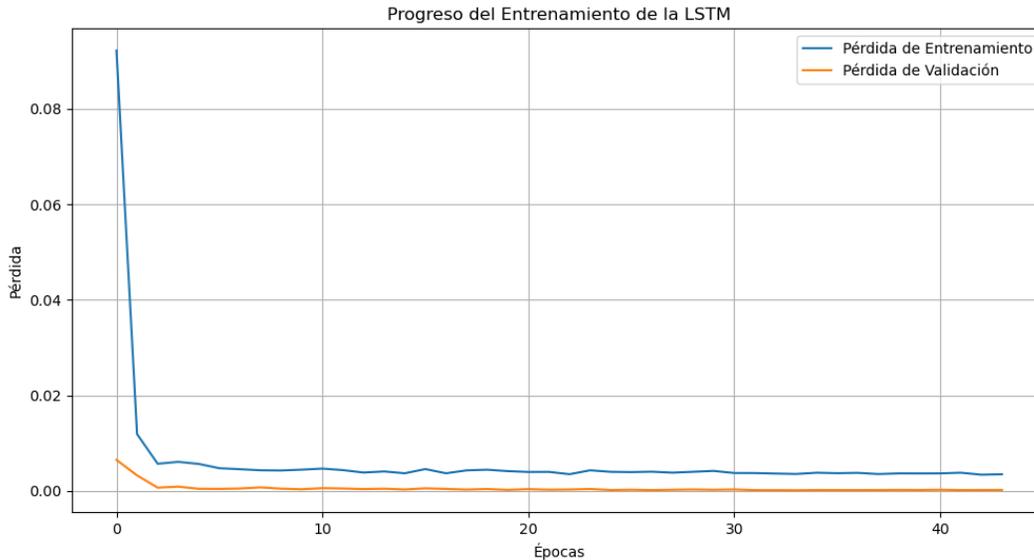


Figura 72: Progreso del entrenamiento del modelo LSTM

En la Figura 72 se ilustra el progreso del entrenamiento del modelo base, mostrando una disminución rápida de la pérdida en las primeras épocas, seguida de una estabilización. Este comportamiento indica que el modelo logró converger de manera eficiente, aunque existieron diferencias entre las pérdidas de entrenamiento y validación, lo que refleja su capacidad limitada para generalizar completamente.

Para mejorar el desempeño del modelo base, se realizó una optimización de hiperparámetros utilizando Keras Tuner, una herramienta que facilita la búsqueda de configuraciones óptimas para arquitecturas neuronales. Este proceso permitió explorar diferentes combinaciones de parámetros clave, como el número de unidades en cada capa LSTM, las tasas de Dropout y la tasa de aprendizaje. Tras completar 30 iteraciones de búsqueda, los mejores hiperparámetros identificados fueron:

- **Unidades LSTM capa 1:** 30
- **Dropout capa 1:** 0.0
- **Unidades LSTM capa 2:** 90
- **Dropout capa 2:** 0.1
- **Tasa de aprendizaje:** 0.0075

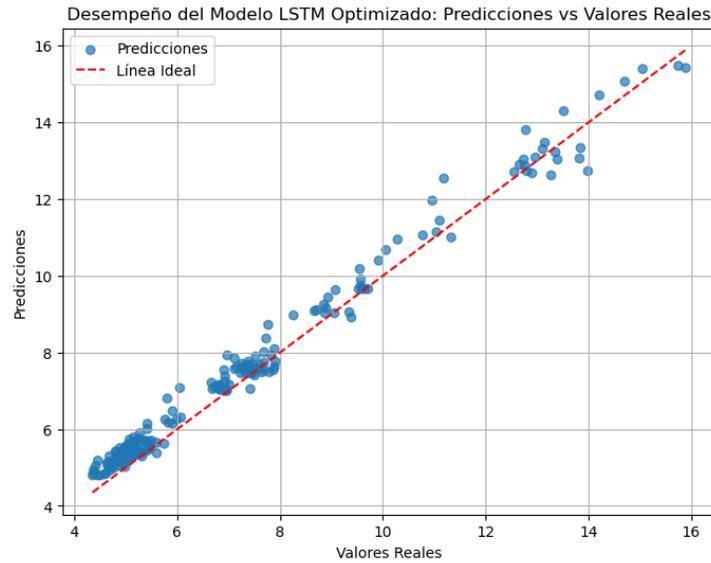


Figura 73: Desempeño del modelo LSTM optimizado: predicciones vs valores reales

El modelo optimizado fue entrenado bajo las mismas condiciones que el modelo base, pero con los hiperparámetros ajustados. Como se muestra en la Figura 73, las predicciones del modelo optimizado están notablemente más alineadas con los valores reales, lo que demuestra una mejora significativa en la precisión. Las métricas cuantitativas del modelo optimizado fueron:

- **Root Mean Squared Error (RMSE):** 0.4408
- **R²:** 0.9757

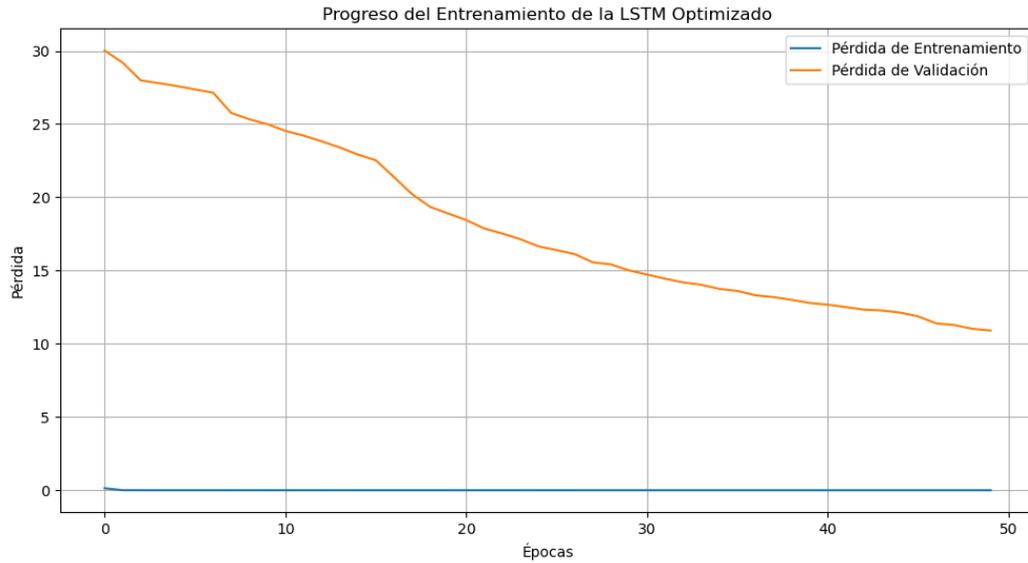


Figura 74: Progreso del entrenamiento del modelo LSTM optimizado

La Figura 74 muestra el progreso del entrenamiento del modelo optimizado. En contraste con el modelo base, se observa una mayor consistencia entre las pérdidas de entrenamiento y validación, lo que indica una mejor capacidad de generalización. Además, el modelo optimizado logró capturar las fluctuaciones del precio del carbono con mayor fidelidad.

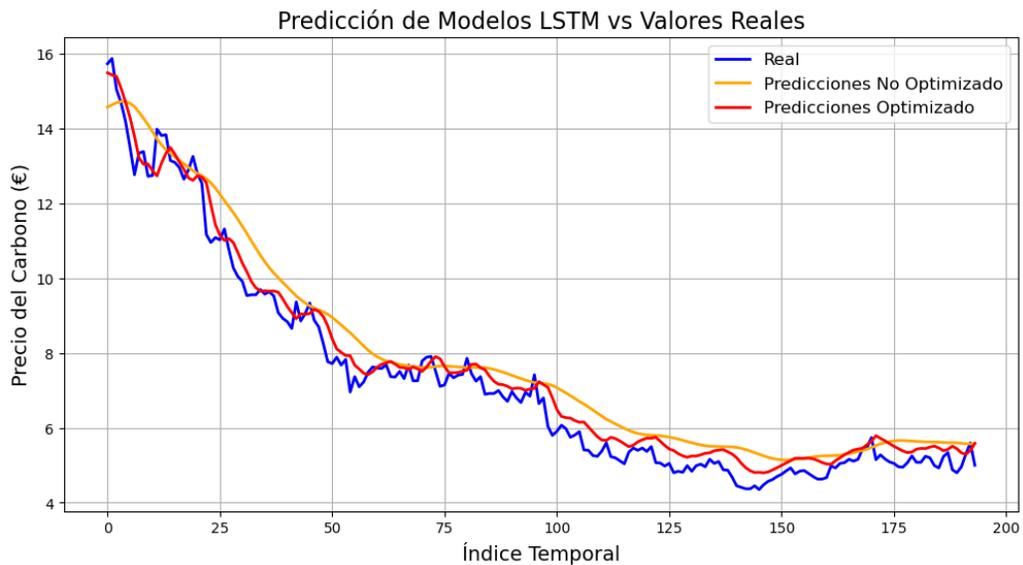


Figura 75: Predicción de modelos LSTM vs valores reales

En la Figura 75, se presenta una comparación visual entre las predicciones del modelo base, el modelo optimizado y los valores reales. Es evidente que el modelo optimizado supera al modelo base al seguir de manera más precisa las variaciones en los precios del carbono, tanto en períodos de alta como de baja volatilidad. Esta comparación destaca la importancia de la optimización de hiperparámetros y su impacto en el desempeño del modelo.

El modelo LSTM optimizado demostró ser una herramienta altamente efectiva para la predicción de precios del carbono. Su capacidad para capturar relaciones temporales complejas y realizar predicciones precisas lo posiciona como una solución robusta para problemas de series temporales. La reducción en el RMSE y el aumento en el coeficiente R^2 validan su desempeño superior en comparación con el modelo base. Este modelo destaca especialmente en escenarios donde las relaciones temporales son cruciales, proporcionando predicciones confiables y detalladas que pueden ser de gran utilidad para la toma de decisiones estratégicas en mercados dinámicos como el del carbono. La optimización de las LSTM permitió desarrollar un modelo que no solo satisface, sino que supera ampliamente las expectativas en cuanto a precisión y capacidad de generalización. Esto demuestra el potencial de las LSTM para aplicaciones en análisis temporal y su relevancia en problemas de predicción de alto impacto como este.

6.10. Comparativa de los Modelos

La comparación de enfoques predictivos resulta esencial para determinar qué método ofrece mejores resultados en términos de precisión, robustez y capacidad de generalización. Dado que el precio del carbono en el EU ETS presenta alta volatilidad y está influido por múltiples factores económicos, energéticos y regulatorios, se han analizado tanto métodos lineales como enfoques no lineales más avanzados, incluyendo redes neuronales especializadas en el tratamiento de series temporales. A continuación, se presenta una discusión detallada de cada categoría de modelos, enfatizando sus fortalezas, limitaciones y el rendimiento alcanzado al aplicarlos a los datos disponibles. Luego, se incluye una tabla que resume la posición de cada modelo según su Error Cuadrático Medio de la Raíz (RMSE).

6.10.1. Enfoques lineales y polinómicos

Los primeros métodos examinados fueron los lineales tradicionales, como la Regresión Lineal Múltiple y las variantes Lasso y Ridge, que añaden regularización para mitigar el sobreajuste. También se incorporó la Regresión Polinómica, la cual intenta capturar no linealidades al aumentar el grado del polinomio de las variables predictoras. Si bien las regresiones lineales ofrecen un marco interpretativo claro y fácil de implementar, las relaciones complejas y la volatilidad de los datos del precio del carbono suelen causar un ajuste deficiente, algo evidente en los altos valores de RMSE y en los coeficientes de determinación (R^2) negativos.

La **Regresión Lineal Múltiple** partía de la premisa de que la variable dependiente puede expresarse como una combinación lineal de los predictores. Si bien es un modelo de rápida implementación y muy utilizado en contextos simples, ha demostrado en este caso ser insuficiente. **El RMSE de 19.7604 y el R^2 de -33.7906** la ubican en la última posición de la tabla, reflejando un ajuste tan pobre que ni siquiera supera a un modelo trivial basado en la media de los valores observados. Esto sugiere que la linealidad impuesta no es capaz de explicar la complejidad ni las fluctuaciones temporales presentes en el conjunto de datos.

Para mejorar la capacidad de capturar relaciones más complejas, se probó la **Regresión Polinómica hasta el grado 6**. Se buscaba así modelar la posible curvatura en la relación entre predictores y variable objetivo. Sin embargo, **el RMSE de 18.0625 y el R^2 de -28.0689** continúan siendo demasiado débiles para un problema tan volátil. El grado 6 resultó ser el óptimo tras diversas pruebas, pero incluso en ese caso, el rendimiento es muy limitado y el modelo no supera la mayoría de los métodos más avanzados.

Las **técnicas de regularización Lasso y Ridge** introducen penalizaciones L1 y L2, respectivamente. En el caso de Lasso, algunos coeficientes pueden anularse por completo, facilitando la selección de variables; en Ridge, se reduce la magnitud de los coeficientes, pero casi nunca llegan a cero. Ambas variantes, tanto en su versión inicial como optimizada, mostraron un incremento notable en la precisión respecto a la Regresión Lineal Múltiple, pero aun así no lograron ofrecer un R^2 positivo. **Lasso optimizado alcanzó un RMSE de 5.7541, mientras que Ridge optimizado llegó a 5.5331**. A pesar de ser mejores que los valores de la regresión pura y la polinómica, no son suficientes para explicar la variabilidad de los datos cuando se observan las fluctuaciones diarias del precio del carbono. El coeficiente de

determinación en estos modelos sigue siendo negativo, lo que reafirma la idea de que un marco lineal (o polinómico clásico) resulta inadecuado para capturar la complejidad inherente a la volatilidad de este mercado.

6.10.2. Modelos basados en árboles de decisión y métodos de ensamble

Con el objetivo de mejorar la capacidad de generalización y capturar relaciones no lineales, se emplearon modelos basados en árboles de decisión y sus extensiones mediante métodos de ensamble. Estos algoritmos suelen ser más flexibles que las regresiones lineales, ya que pueden particionar el espacio de predictores de manera recursiva y jerárquica, dando lugar a reglas que se ajustan a la forma de los datos. Sin embargo, su eficacia depende en gran medida de la optimización de hiperparámetros como la profundidad del árbol, el número de estimadores y la tasa de aprendizaje, entre otros.

El **Árbol de Decisión**, en su versión de regresión, mostró limitaciones notorias para un problema tan volátil. En la configuración inicial, **tuvo un RMSE de 16.5795 y un R^2 de -23.4914, cifras que apenas mejoraron a 16.1807 y -22.3272 tras optimizar parámetros** como la profundidad máxima o el número mínimo de muestras por nodo. Aunque los árboles de decisión ofrecen interpretabilidad y facilidad de uso, la alta varianza de los datos y la necesidad de capturar factores no lineales complejos superan la capacidad predictiva de este modelo individual.

Como alternativa más robusta, se aplicaron métodos de ensamble. **Random Forest** consiste en entrenar múltiples árboles de decisión con datos muestreados aleatoriamente. Al promediar las predicciones de dichos árboles, se reduce la varianza y se obtiene mayor estabilidad que un árbol único. Aun así, los resultados iniciales de este enfoque **evidenciaron un RMSE de 14.7101 y un R^2 de -18.2798**, mejor que un solo árbol pero sin llegar a niveles adecuados de precisión. **La versión optimizada obtuvo 14.4227 y -17.5337**, una mejora marginal que confirma la complejidad de los datos y la necesidad de seguir buscando técnicas más potentes para este tipo de predicción.

El **Gradient Boosting (GBM)** aprovecha un enfoque de boosting, donde los árboles se construyen de forma secuencial para corregir los errores de los modelos anteriores. Esto suele conferirle cierta ventaja sobre el bagging de Random Forest, al focalizarse en aquellas regiones

del espacio de datos donde el modelo actual está fallando. **La versión inicial de GBM arrojó un RMSE de 12.9967, con un R^2 de -14.0500. Al optimizar hiperparámetros clave como la tasa de aprendizaje, la profundidad de los árboles y el número de estimadores, mejoró a 12.1561 en el RMSE y -12.1663 en el R^2 .** Aunque la reducción de errores frente a Random Forest y Árbol de Decisión es apreciable, el coeficiente de determinación continúa siendo negativo.

XGBoost (Extreme Gradient Boosting) se presenta como una versión avanzada de GBM que integra optimizaciones de memoria y de gradientes de segundo orden, además de regularizaciones L1 y L2, lo cual reduce la posibilidad de sobreajuste. **Con parámetros iniciales por defecto, alcanzó un RMSE de 15.5713 y un R^2 de -20.6032. Al optimizar el número de estimadores, la profundidad máxima, la tasa de aprendizaje y la fracción de muestras y de características, se redujo el RMSE hasta 11.6919.** Este valor es notablemente mejor que el de GBM y Random Forest, aunque el R^2 permanece negativo (-11.1799). Por tanto, XGBoost optimizado se ubica como el modelo de ensamble que consigue el menor RMSE, si bien no logra competir con otras arquitecturas más apropiadas para series temporales complejas.

6.10.3. Redes Neuronales (MLP y LSTM)

El uso de redes neuronales se justifica por la habilidad que tienen para capturar relaciones no lineales y, en el caso de las recurrentes, por su capacidad de procesar datos secuenciales. El mercado del carbono, con su alta volatilidad y su naturaleza dependiente del tiempo, se ajusta con frecuencia a este tipo de metodologías, especialmente cuando se cuenta con la información adecuada para entrenar.

El **Multilayer Perceptron (MLP)**, en su configuración inicial, **mostró un RMSE de 15.19 y un R^2 de -19.54, valores apenas superiores a algunos métodos de ensamble.** Sin embargo, tras un ajuste sistemático de hiperparámetros que incluyó el número de neuronas, las funciones de activación, la tasa de aprendizaje y el uso de regularización por dropout, el modelo mejoró significativamente. **El MLP optimizado alcanzó 6.80 de RMSE, con un R^2 de -3.11.** Aun siendo negativo, esta reducción de error lo posiciona muy por encima de la mayoría de los métodos lineales y de algunos ensambles, lo que demuestra la importancia de una sintonización

profunda y la capacidad del MLP para modelar no linealidades en un contexto no secuencial estricto.

La **Red Neuronal Long Short-Term Memory (LSTM)** se diseñó de forma específica para manejar series temporales y retener información histórica a más largo plazo. A través de celdas de memoria y puertas de control (entrada, olvido y salida), las LSTM logran atenuar el desvanecimiento de gradientes, un problema frecuente en redes recurrentes tradicionales. **La implementación inicial del modelo LSTM ya mostró resultados prometedores, con un RMSE de 1.0262 y un R^2 de 0.8686.** Estas cifras suben el nivel de precisión por encima de casi todos los modelos basados en árboles y de las regresiones lineales, evidenciando la ventaja de procesar explícitamente la estructura secuencial de los datos en un mercado tan dinámico.

Sin embargo, **la verdadera mejora se observó al optimizar la arquitectura LSTM.** Al ajustar parámetros como el número de unidades LSTM en cada capa, la tasa de aprendizaje, el grado de dropout y el número de épocas de entrenamiento, **el RMSE descendió drásticamente a 0.4408 y el R^2 alcanzó 0.9757, situando al LSTM optimizado en la primera posición de la clasificación.** Estos valores reflejan que el modelo no solo minimiza el error en la escala absoluta, sino que también explica la mayor parte de la variabilidad observada en la serie temporal. Tal nivel de desempeño se traduce en predicciones de alta fiabilidad, especialmente si se toma en cuenta que los datos representan un mercado real con diversos factores externos en juego.

En la Figura 76, se aprecia la evolución temporal tanto del precio real del carbono (trazo negro) como de las predicciones obtenidas por cada modelo (colores diferenciados). En el eje vertical se representa el precio en euros, mientras que el eje horizontal muestra el índice temporal (o los pasos de la serie).

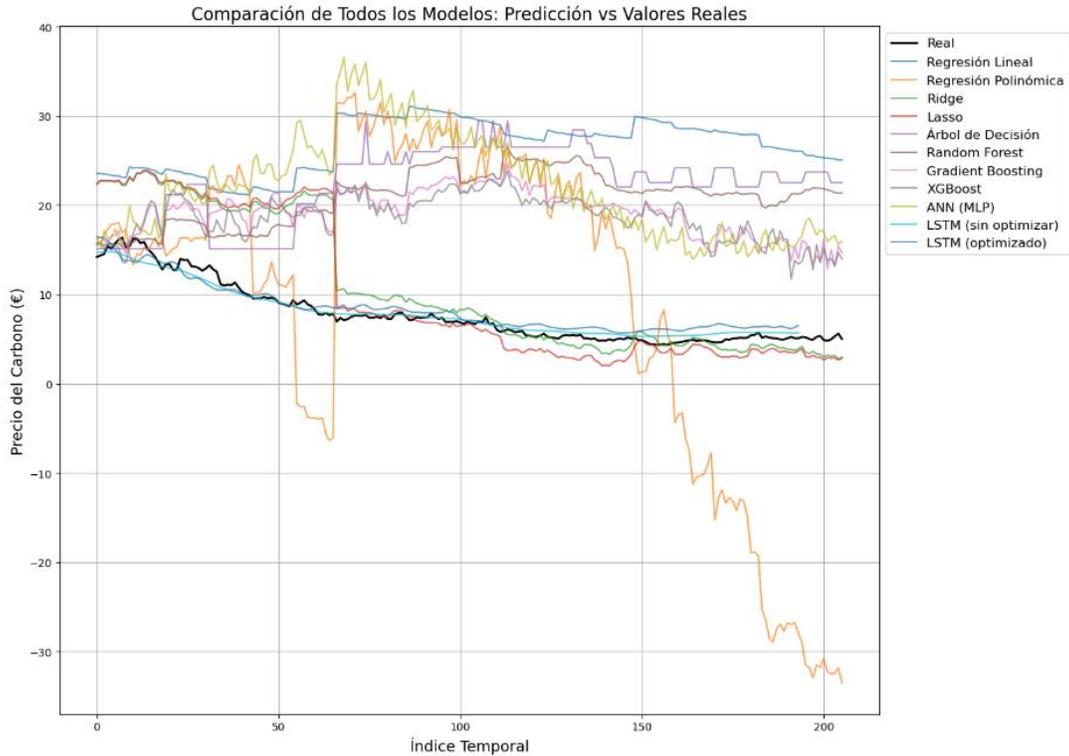


Figura 76: Comparación de todos los modelos: predicción vs. valores reales

Tal y como se observa, la mayoría de los modelos lineales (Regresión Lineal, Regresión Polinómica, Ridge, Lasso) presentan una subestimación o sobreestimación continuada, incapaces de seguir con fidelidad los cambios bruscos de la serie. El Árbol de Decisión y sus variaciones en ensamble (Random Forest, Gradient Boosting, XGBoost) parecen reaccionar un poco mejor a ciertas oscilaciones, pero en general se mantienen separados de la verdadera trayectoria del precio.

Llama la atención la línea correspondiente a la Red Neuronal LSTM, especialmente en su versión optimizada, que se muestra muy cercana al trazo negro (el valor real) a lo largo de gran parte del horizonte temporal. De forma similar, el MLP optimizado ofrece un ajuste razonablemente decente si se compara con otras líneas, pero todavía no alcanza la precisión de la LSTM.

En la zona central de la gráfica puede observarse cómo algunos modelos sufren altos picos o caídas excesivas, mientras que el precio real mantiene una tendencia menos extrema. Esto pone

en evidencia la insuficiente capacidad de las aproximaciones lineales o de los árboles, incluso en sus versiones mejoradas. La LSTM, por su parte, consigue adaptarse a transiciones relativamente rápidas y a comportamientos estacionarios, reflejando así la fortaleza de las celdas de memoria para codificar patrones de largo plazo.

La siguiente tabla muestra la clasificación de los modelos en función de su RMSE, desde el más bajo (mejor desempeño) hasta el más elevado (peor ajuste). También se incluyen los valores de R^2 para cada modelo.

Posición	Modelo	RMSE	R^2
1	LSTM (Optimizado)	0.4408	0.9757
2	LSTM (Inicial)	1.0262	0.8686
3	Ridge (Optimizado)	5.5331	-1.7278
4	Lasso (Optimizado)	5.7541	-1.9500
5	Lasso (Inicial)	5.8670	-2.0670
6	Ridge (Inicial)	6.1887	-2.4125
7	MLP (Optimizado)	6.80	-3.11
8	XGBoost (Optimizado)	11.6919	-11.1799
9	Gradient Boosting (Optimizado)	12.1561	-12.1663
10	Gradient Boosting (Inicial)	12.9967	-14.0500
11	Random Forest (Optimizado)	14.4227	-17.5337

12	Random Forest (Inicial)	14.7101	-18.2798
13	MLP (Inicial)	15.19	-19.54
14	XGBoost (Inicial)	15.5713	-20.6032
15	Árbol de Decisión (Optimizado)	16.1807	-22.3272
16	Árbol de Decisión (Inicial)	16.5795	-23.4914
17	Regresión Polinómica (Grado 6)	18.0625	-28.0689
18	Regresión Lineal Múltiple	19.7604	-33.7906

Tabla 9: Tabla comparativa de los resultados de los modelos

Tal como puede observarse, la red LSTM en su versión optimizada alcanza el primer lugar, seguida por la LSTM inicial y, mucho más lejos, por algunos modelos lineales regularizados, como Ridge y Lasso optimizados. El resto de métodos, principalmente los árboles y sus ensambles, presentan RMSE más elevados y valores de R^2 negativos.

7. Conclusión

El análisis confirma que los métodos lineales (Regresión Lineal, Polinómica, Lasso, Ridge) y los basados en árboles (Árbol de Decisión, Random Forest, Gradient Boosting, XGBoost) no alcanzan la precisión necesaria para describir un mercado tan complejo como el EU ETS. La persistencia de valores negativos de R^2 , junto con RMSE elevados, recalca la dificultad de representar patrones subyacentes con enfoques que no aprovechan la naturaleza secuencial de los datos.

Las redes neuronales, en particular la LSTM optimizada, muestran un desempeño muy superior. Al incorporar la dimensión temporal y permitir la propagación de información relevante a lo largo de la serie, la LSTM logra capturar fluctuaciones que otros modelos ignoran. Esta ventaja se hace especialmente notoria ante picos, caídas repentinas o cambios de tendencia en el precio del carbono, como se aprecia claramente en la figura comparativa.

Sin embargo, el empleo de redes neuronales conlleva ciertos desafíos. Por un lado, el coste computacional y el tiempo de entrenamiento resultan mayores que en la mayoría de los métodos precedentes, por lo que la infraestructura de cómputo y la eficiencia del proceso de entrenamiento se vuelven aspectos críticos. Por otro lado, las redes neuronales suelen considerarse “cajas negras” desde la perspectiva de interpretabilidad, lo que puede suponer una barrera cuando se requiere explicar con claridad la base de una decisión financiera o regulatoria.

Aun así, el potencial de LSTM queda avalado por estos resultados: no solo obtiene un RMSE en torno a 0.44, sino que presenta un R^2 de 0.9757, lo cual implica una capacidad explicativa muy elevada respecto a la serie de datos analizada. Para entornos productivos, la adopción de una LSTM o de redes neuronales recurrentes similares (como GRU o variantes con mecanismos de atención) facilitaría la toma de decisiones informadas, la gestión de riesgo y la planificación estratégica del uso de derechos de emisión, ya que permitiría anticipar con mayor fiabilidad los movimientos del mercado del carbono.

En síntesis, la comparación efectuada, apoyada tanto en la tabla de resultados como en la figura de predicción vs. valores reales, pone de manifiesto que la LSTM optimizada destaca de forma incuestionable frente al resto de métodos. La naturaleza no lineal, el componente temporal y la volatilidad del precio del carbono confluyen para hacer especialmente ventajoso el uso de redes neuronales recurrentes con celdas de memoria de largo plazo. Mientras que otros enfoques se quedan cortos, bien por su linealidad implícita o por la falta de capacidad para “recordar” la evolución de la serie, la LSTM confirma ser la alternativa más robusta y precisa para este tipo de predicción.

8. Bibliografía

- [1] D. Ellerman, F. J. Convery, and C. de Perthuis. Pricing Carbon: The European Union Emissions Trading Scheme. Cambridge University Press, 2010.
- [2] Organisation for Economic Co-operation and Development (OECD). Carbon Pricing: What Role for Border Carbon Adjustments. Technical Report, 2021.
- [3] Journal of Cleaner Production. The Impact of Carbon Price and Firm-Level Emissions on Swedish Firms' Profits. 2023.
- [4] Energy Economics. The Environmental-Financial Performance Nexus of EU ETS Firms: A Quantile Regression Approach. 2022.
- [5] K. S. Rogge, M. Schneider, and V. H. Hoffmann. The Innovation Impact of the EU Emission Trading System: Findings of Company Case Studies in the German Power Sector. Ecological Economics, 70(2011), 513–523.
- [6] BI Norwegian Business School. EU ETS & the Maritime Shipping Industry: Barriers Compromising the Economic Efficiency and Effectiveness of the System. 2024.
- [7] Cambridge Press. The European Union's Emissions Trading System in Perspective. 2019.
- [8] R. Rafaty, G. Dolphin, and F. Pretis. Carbon Pricing and the Elasticity of CO₂ Emissions. Energy Policy Research Group, University of Cambridge, Working Paper 2035, 2020.
- [9] A. Wach. The European Union Emissions Trading Scheme: Failure Analysis and Assessment of Market Stability Reserve Solution. Norwegian School of Economics, Bergen, 2016.
- [10] European Commission. The EU Emissions Trading System. European Commission, 2021.
- [11] L. H. Goulder and A. R. Schein. Carbon Pricing in the EU: Evaluation of Different EU ETS Reform Options. Technical Report, 2022.
- [12] C. Flachsland, N. Koch, and S. Fuss. A Framework for Assessing the Performance of Cap-and-Trade Systems: Insights from the EU ETS. Review of Environmental Economics and Policy, 2018.
- [13] M. A. Mehling and R. A. Ritz. Emissions Trading Hybrids: The Case of the EU ETS. Energy Economics, 2022.
- [14] European Commission, Climate Action, EU ETS Emissions Cap. [En línea]. Disponible en: https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/eu-ets-emissions-cap_en.
- [15] M. Cames and M. Frerk. The Best (and Worst) of GHG Emission Trading Systems: Comparing the EU ETS with Its Followers. Journal of Environmental Policy, 2019.

- [16] R. Martin, M. Muûls, and U. J. Wagner. Allocation and Competitiveness in the EU Emissions Trading Scheme: Policy Overview. Carbon Markets Series, 2020.
- [17] F. Laurent. The European Ships' Monitoring, Reporting and Verification (MRV): Pre-evaluation of a Regional Regulation on Carbon Dioxide Inventory. IAME Conference, 2016.
- [18] J. Verschuuren and F. Fleurke. Report on the Legal Implementation of the EU ETS at Member State Level. Tilburg University, 2014.
- [19] J. Kruger and C. Egenhofer. Confidence through Compliance in Emissions Trading Markets. Sustainable Development Law & Policy, 2005.
- [20] T. Santier. Monitoring, Reporting, and Verification in a Carbon Border Adjustment Mechanism. 2023.
- [21] M. Zhang, Y. Liu, and Y. Su. Comparison of Carbon Emission Trading Schemes in the European Union and China. Climate, 2017.
- [22] P. Gailhofer and V. Graichen. Monitoring, Reporting and Verification in the EU ETS: Legal Design Considerations. 2024.
- [23] C. Warnecke. Can CDM Monitoring Requirements Be Reduced While Maintaining Environmental Integrity?. Climate Policy, 2014.
- [24] F. Deane, A. Huggins, and M. S. Karim. Measuring, Monitoring, Reporting and Verification of Shipping Emissions. Review of European, Comparative & International Environmental Law, 2019.
- [25] K. Kulovesi, E. Mekkonen, y T. Voigt. Enforcement of the EU ETS in the Member States. Journal of Environmental Law, 2021.
- [26] Red Nacional de Derechos de Emisión (RENADE), Taxes Renade. [En línea]. Disponible en: <https://www.renade.es/ing/Information/Taxes-Renade>.
- [27] J. Schleich, R. Betz, y K. Rogge. EU Emissions Trading – Better Job Second Time Around. European Environment, 2007.
- [28] J. Schleich, R. Betz, and K. Rogge. EU Emissions Trading – Better Job Second Time Around?. European Environment, 17(5), 279–294, 2007.
- [29] J. Wettestad and T. Jevnaker. Rescuing EU Emissions Trading: The Climate Policy Flagship. Springer, 2016.
- [30] J. B. Skjærseth and J. Wettestad. EU Emissions Trading: Initiation, Decision-Making and Implementation. Global Environmental Politics, 8(1), 117–141, 2008.
- [31] European Commission. The EU Emissions Trading System (EU ETS). Technical Report, European Commission, 2021.
- [32] G. Perino y M. Willner. Procrastinating Reform: The Impact of the Market Stability Reserve on the EU ETS. Journal of Environmental Economics and Management, 2016.

- [33] C. Flachsland, R. Marschinski, and O. Edenhofer. To Link or Not to Link: Benefits and Disadvantages of Linking Cap-and-Trade Systems. *Climate Policy*, 9(4), 358–372, 2009.
- [34] J. Köhler, M. Grubb, and D. Popp. Transition Pathways and Risk Analysis for Climate Change Mitigation. *Climate Policy*, 19(3), 342–358, 2019.
- [35] M. Duwe, J. Graichen, y H. Böttcher. Can Current EU Climate Policy Reliably Achieve Climate Neutrality by 2050?. Ecologic Institute, 2023.
- [36] J. Boratyński, M. Pyrka, S. Lizak, I. Tobiasz, y R. Jeszke. Reform of the Market Stability Reserve (MSR) in the “Fit for 55” Package. IOS-PIB, 2022.
- [37] M. Wiels and J. C. Defraigne. The EU's Climate Ambitions: Assessment of the Social Cost of the Fit for 55 Package. Université Catholique de Louvain, 2023.
- [38] J. C. Richstein, É. J. L. Chappin, y L. J. de Vries. The Market (In-)Stability Reserve for EU Carbon Emission Trading: Why It Might Fail and How to Improve It. *Utilities Policy*, 2015.
- [39] D. Seibert, P. Kasten, J. Graichen, y N. Wissner. EU 2040 Climate Target: Contributions of the Transport Sector. Oeko Institute, 2024.
- [40] H. Rienks y B. Moore. EU Innovation Funding for Climate Neutrality. Wageningen University, 2023.
- [41] T. Hesse y S. Braungardt. EU 2040 Climate Target: Contributions of the Buildings Sector. Oeko Institute, 2024.
- [42] B. Chen and R. Wu. Comparative Analysis of the Carbon Market in the EU and China. *Journal of Environmental Policy*, 2023.
- [43] F. Annunziata. Emission Allowances as Financial Instruments. In *Governance, Financial Stability and Financial Markets*, Springer, 2021.
- [44] A. Michaelowa and I. Shishlov. Evolution of International Carbon Markets: Lessons for the Paris Agreement. *Wiley Interdisciplinary Reviews: Climate Change*, 10(6), e613, 2019.
- [45] M. Li and M. Duan. Exploring Linkage Opportunities for China's Emissions Trading System Under the Paris Targets: EU-China and Japan-Korea-China Cases. *Energy Economics*, 2021.
- [46] World Bank, Carbon Pricing Dashboard. [En línea]. Disponible en: <https://carbonpricingdashboard.worldbank.org/>.
- [47] C. Müller. Rethinking Voluntary Carbon Credit Markets and the EU ETS: A Critical Review. Fulda University of Applied Sciences, 2024.
- [48] D. Tänzler, J. Groß, and C. Warnecke. Interactions Between New Market Mechanisms and Emissions Trading Schemes: Opportunities and Prospects for Countries to Use Article 6 of the Paris Agreement. German Federal Ministry for the Environment, 2018.

- [49] O. M. Hladka, I. M. Karpovich, and A. Y. Sabaniuk. Artificial Neural Networks for Analysis of Stock Market Dynamics: Applications of Python and Jupyter Notebooks. *Vernadsky Tech Journals*, 2024.
- [50] E. Champa-Bujaico, A. M. Díez-Pascual, A. Lomas Redondo, and P. Garcia-Diaz. Optimization of mechanical properties of multiscale hybrid polymer nanocomposites: A combination of experimental and machine learning techniques. *Composites Part B*, vol. 269, p. 111099, 2024.
- [51] R. Bhattacharjee, K. Botchway, J. C. Pashin, G. Chakraborty, and P. Bikkina, "Machine learning-based prediction of CO₂ fugacity coefficients: Application to estimation of CO₂ solubility in aqueous brines as a function of pressure, temperature, and salinity," *International Journal of Greenhouse Gas Control*, vol. 128, p. 103971, 2023.
- [52] T. Souza, J. Herrera, A. Johnson, and K. B. Wright. Detection and classification of pneumonia using the Orange3 data mining tool. *International Journal of Computational Intelligence Systems*, vol. 14, no. 3, pp. 250-263, 2023.
- [53] F. Cheng, C. Yang, C. Zhou, L. Lan, H. Zhu, and Y. Li. Simultaneous Determination of Metal Ions in Zinc Sulfate Solution Using UV–Vis Spectrometry and SPSE-XGBoost Method. *Sensors*, vol. 20, no. 17, p. 4936, 2020.
- [54] K. Y. Chan, B. Abu-Salih, R. Qaddoura, A. M. Al-Zoubi, V. Palade, D.-S. Pham, J. Del Ser, and K. Muhammad. "Deep neural networks in the cloud: Review, applications, challenges and research directions." *Neurocomputing*, vol. 545, p. 126327, 2023.
- [55] B. Yang, Z. Guo, L. Wang, J. He, B. Xia, and S. Vakily. Updated Global Navigation Satellite System Observations and Attention-Based Convolutional Neural Network–Long Short-Term Memory Network Deep Learning Algorithms to Predict Landslide Spatiotemporal Displacement. *Remote Sens.*, vol. 15, no. 20, p. 4971, 2023.

Anexo I

Alineación con los Objetivos de Desarrollo Sostenible (ODS):

Este proyecto de fin de máster se alinea con varios Objetivos de Desarrollo Sostenible (ODS) establecidos por las Naciones Unidas, contribuyendo de manera directa e indirecta a su cumplimiento. Su enfoque en la sostenibilidad y en la predicción de dinámicas clave del mercado de emisiones refuerza la conexión con las metas globales. A continuación, se detallan los ODS con los que el proyecto tiene mayor sinergia:

1. ODS 7: Energía asequible y no contaminante.

El análisis y la predicción de la dinámica del mercado de derechos de emisión pueden contribuir a incentivar la inversión en tecnologías limpias y energías renovables, reduciendo la dependencia de combustibles fósiles y promoviendo el uso de energía asequible, fiable, sostenible y moderna. Además, al optimizar la gestión de emisiones, fomenta un entorno más favorable para el desarrollo de soluciones energéticas innovadoras y accesibles.

2. ODS 9: Industria, innovación e infraestructura.

Mediante el desarrollo de modelos predictivos y técnicas de análisis avanzadas, este proyecto fomenta la innovación tecnológica en el sector industrial. Facilita la adaptación y creación de infraestructuras resilientes y sostenibles que minimicen el impacto ambiental de la producción y los servicios. El uso de herramientas avanzadas también impulsa la capacidad de las industrias para responder a los desafíos del cambio climático de manera ágil y eficiente.

3. ODS 12: Producción y consumo responsables.

Al predecir cambios en el mercado de derechos de emisión y entender mejor cómo las prácticas empresariales afectan este mercado, el proyecto promueve patrones de producción y consumo más sostenibles. Ayuda a las empresas a adoptar prácticas más eficientes y menos contaminantes, alineando sus operaciones con principios de sostenibilidad. Además, fomenta

una mayor conciencia sobre la relación entre las políticas ambientales y las decisiones de consumo.

4. ODS 13: Acción por el clima.

Este es el ODS con el que el proyecto se alinea de forma más directa. Al enfocarse en el Sistema de Comercio de Emisiones de la Unión Europea (EU ETS) y buscar formas más efectivas de predecir y gestionar las emisiones de gases de efecto invernadero, el proyecto contribuye significativamente a la lucha contra el cambio climático. Facilita la toma de decisiones basadas en evidencia para la implementación de políticas climáticas efectivas y la reducción de las emisiones globales. Asimismo, refuerza la cooperación internacional al ofrecer herramientas útiles para la planificación de estrategias climáticas conjuntas.