Keeping Prices Cool:
How Machine Learning Transforms Refrigerator Pricing Strategies

Tudor GHINEA*
Bucharest University of Economic Studies, Bucharest, Romania
**Corresponding author, ghineatudor22@stud.ase.ro*

Răzvan-Gabriel FIRAN
Bucharest University of Economic Studies, Bucharest, Romania
firanrazvan22@stud.ase.ro

Cristian-Maximilian GOGA
Bucharest University of Economic Studies, Bucharest, Romania
gogacristian22@stud.ase.ro

Marina-Diana AGAFIȚEI
Bucharest University of Economic Studies/National Scientific Research Institute for Labour and Social Protection, Bucharest, Romania
diana.agafiei@csie.ase.ro

Abstract. *The demand prediction of product pricing is in retail and a key challenge to balancing profitability and market competitiveness. This study explores the potential of machine learning algorithms for predicting refrigerator prices in the household appliance industry, filling a gap between classical statistical approaches and sophisticated regression models. Existing literature reflects a trend towards algorithmic price modelling, yet few specifically focus on the consumer durables sector, with studies observing mostly real estate or commodities rather than differentiated retail products. The research uses eight different machine learning methods to analyze a rich dataset of refrigerators from tree-based ensemble methods to support vector machines and neural networks. The models use the various features of a product, including technical specifications (volume, energy efficiency), aesthetics (colour, design), and brand positioning to predict market prices. Results show that ensemble models have a significantly better prediction power than traditional approaches, with the Random Forest algorithm achieving significantly lower error metrics than simple regression models. Clustering analysis shows that algorithms have different performance tiers, with tree-based methods being the highest-performing cluster. Finally, our findings provide meaningful implications for the development of retail pricing strategies by identifying economically significant product attributes in various market segments. The study's core contribution is in building a methodological framework for well-defined price prediction in markets with differentiated products that support equity pricing in retail and can be translated across product categories to change the way pricing strategies are set beyond the household appliance industry.*

Keywords: Price Prediction, Machine Learning, Ensemble Methods, E-Commerce, Feature Importance, Refrigerator Market, Clustering Analysis.

Introduction

The rapid growth of the e-commerce market for household appliances has created a complex pricing landscape where retailers must balance competitiveness with profitability.

DOI: 10.2478/picbe-2025-0203

Refrigerators represent a particularly challenging product category due to their essential nature, significant price range, and a multitude of technical specifications that influence consumer purchasing decisions. Traditional pricing strategies often rely on simplistic cost-plus approaches or competitive benchmarking, failing to capture the intricate relationships between product attributes and perceived value. As e-commerce platforms accumulate vast amounts of product data, including technical specifications, customer reviews, and sales performance, there exists an unprecedented opportunity to leverage advanced analytical techniques for more sophisticated price modelling. Recent research indicates that consumers increasingly consider factors beyond basic functionality when purchasing refrigerators, with energy efficiency, design aesthetics, and brand reputation significantly influencing willingness to pay. Additionally, market segmentation has created distinct consumer profiles with varying price sensitivities to specific product attributes. This complexity makes refrigerator pricing an ideal candidate for machine learning approaches that can identify non-linear relationships and complex interactions between variables that traditional statistical methods might overlook.

This study aims to evaluate the comparative performance of eight machine learning algorithms, including Random Forest, Support Vector Machines, XGBoost, and Neural Networks, in predicting refrigerator prices based on technical specifications and design attributes.

The research contributes to the domain through the development of a clustering-based model evaluation methodology that provides a more robust algorithm selection compared to traditional single-metric approaches, representing a methodological innovation in predictive modelling for retail pricing. Furthermore, this study introduces segment-specific error analysis that reveals how prediction accuracy varies across price ranges, offering practical insights into market segmentation and pricing strategy development that extend beyond the specific case of refrigerators to broader retail applications.

The paper is organized as follows: first, we present a comprehensive literature review of price prediction methods and machine learning applications in retail; next, we describe our methodology involving the selected machine learning models applied to refrigerator pricing data; then, we analyze the results comparing model performance and conducting clustering analysis; finally, we discuss the implications for retailers and manufacturers, acknowledge limitations, and suggest directions for future research.

Literature review

Recent advances in machine learning have significantly enhanced price prediction models across various domains. This study builds upon established methodological frameworks while introducing key refinements to improve predictive performance.

Orji and Ukwandu (2024) demonstrated the efficacy of ensemble methods, particularly Random Forest and XGBoost, for price prediction. Our methodology extends their work through systematic hyperparameter optimization and insights from Asghar et al. (2021), who explored machine learning for real estate price prediction in emerging markets, and Ho et al. (2021), who investigated housing valuation using advanced regression techniques.

Our Random Forest implementation refines Orji and Ukwandu's approach by optimizing the mtry parameter, addressing the bias-variance tradeoff more effectively, as supported by Shahhosseini et al. (2022). Similarly, our XGBoost implementation extends prior work by simultaneously tuning iteration count, tree complexity, and learning rate, mitigating overfitting risks while maximizing predictive capacity, as highlighted by Bentéjac et al. (2021). A significant extension is the incorporation of Support Vector Machines (SVM) with radial basis function kernels, addressing Mitra's (2022) observation that kernel-based methods effectively capture non-linear price determinants. Our SVM implementation emphasizes meticulous feature standardization through Z-score normalization, ensuring optimal hyperplane formation.

In data preprocessing, we employ a robust outlier detection method using the Interquartile Range (IQR) with 1.5×IQR boundaries, addressing limitations in Asghar et al.'s (2021) min-max normalization and Orji and Ukwandu's (2024) outlier handling. This aligns with Ho et al.'s (2020) emphasis on robust preprocessing but extends their approach through more comprehensive outlier handling and feature normalization. Our feature engineering approach surpasses previous studies by deriving composite variables such as volume-to-height ratios and efficiency-per-volume metrics, moving beyond Asghar et al.'s reliance on original features and Ho et al.'s statistical transformations.

A key innovation is the implementation of the Boruta algorithm for feature selection, replacing simpler methods like correlation analysis (Asghar et al., 2023), stepwise selection (Ho et al., 2020), and basic feature importance metrics (Orji & Ukwandu, 2024). This wrapper-based method provides statistically rigorous feature importance assessment, enhancing model parsimony and addressing the "curse of dimensionality" noted by Orji and Ukwandu(2024).

Our methodology is further contextualized by comparison with studies employing different modelling approaches. Deng et al. (2022) explored deep learning techniques, specifically Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN), for cryptocurrency price prediction. While their temporal models excel in capturing sequential dependencies, our ensemble and kernel-based methods achieve competitive results without requiring explicit temporal modelling, aligning with Bentéjac et al.'s (2021) findings on the versatility of gradient boosting frameworks. Unlike Deng et al.'s reliance on lagged variables and technical indicators, our methodology emphasizes feature engineering, creating derived variables that align with Mitra's (2022) observation that data transformation can rival architectural complexity in performance improvements.

M'hamdi et al. (2024) employed Random Forest for plant disease detection, providing a parallel to our use of ensemble methods. However, our implementation extends theirs through comprehensive hyperparameter optimization, including grid searches for mtry (Random Forest) and nrounds, max_depth, and learning rate (XGBoost). This addresses Shahhosseini et al.'s (2022) emphasis on the impact of hyperparameter selection on ensemble model performance. Additionally, our hybrid approach combining Random Forest, XGBoost, and SVM represents a methodological advancement over single-model approaches, leveraging the strengths of diverse algorithms for more robust predictions.

Our 10-fold cross-validation strategy, employing multiple metrics (RMSE, MAE, MAPE, R^2), provides more reliable performance estimates than the basic train-test splits used by Deng et al. (2022) or the limited cross-validation in M'hamdi et al. (2024). This comprehensive validation framework addresses concerns about result stability noted in both studies.

Recent studies such as Zehtab-Salmasi et al. (2020) have also highlighted the importance of integrating multiple data modalities to enhance predictive accuracy. Similarly, Safonov (2024) emphasizes the role of neural networks in capturing complex demand patterns and optimizing pricing strategies. Furthermore Hasnain et al. (2024) demonstrates the efficacy of machine learning models in predicting prices for used electronics, reinforcing the applicability of these techniques across various product categories.

In summary, our methodology integrates strengths from previous studies while addressing their limitations. We maintain the ensemble method focus advocated by Orji and Ukwandu (2024), implement robust preprocessing emphasized by Ho et al. (2021), and extend beyond the basic validation approaches used by Asghar et al. (2021). By incorporating insights from Deng et al. (2022) and M'hamdi et al. (2024), we demonstrate the versatility of ensemble methods and the importance of feature engineering and hyperparameter optimization, offering a comprehensive and refined approach to machine learning-based price prediction.

Data and methodology

This research utilizes a dataset obtained through web scraping from the eMag platform, one of the largest e-commerce platforms in Romania. The dataset has allowed us to have access to a significant volume of information about household appliances. However, this methodology presents several limitations that should be mentioned. Product prices are dynamic and can undergo changes, they can fluctuate daily depending on promotions, available stock, and commercial marketing strategies, so the predictive models developed reflect only a temporal snapshot of the market. The data is not extracted over a period of time, but only at a specific chosen moment. Also, some product characteristics may be missing or inconsistent across different listings, which may affect the accuracy of the analyses performed. For data collection, Python programming language was used together with the BeautifulSoup library, which enabled the automated extraction of information about refrigerators available on the platform. The data was then saved in CSV/Excel format for subsequent processing. The collected dataset contains detailed information about refrigerators, structured across multiple dimensions essential for analyzing prices and technical characteristics. Each record includes a unique identifier (e.g., "id", string type) and the complete product name ("nume", string), which provides descriptive details, including technical and aesthetic specifications. There are also fields such as "imagine_url" (string) and "url_produs" (string), which provide links to the product image and the product presentation page on the eMag platform. From a numerical data perspective, the "pret" (float) variable represents the listing price, and "rating" (float) indicates the average customer ratings, while "nr_reviewuri" (integer) quantifies the number of reviews. These are essential for analyzing correlations between popularity, perceived quality, and price. Additionally, there are variables related to energy consumption, such as "Consum_anual_energie" and "Consum_zilnic_energie" (both float type), which measure energy consumption in corresponding units (e.g., kWh/year and kWh/day). The data also includes dimension-related variables, such as "Inaltime" (integer), and volume can be implicitly calculated, although volume is included in the product name and can be extracted as a numerical variable through subsequent processing.

Regarding technical and aesthetic aspects, the dataset contains binary indicators (0 or 1) for variables such as the number of doors – "O_usa" and "Doua_usi" – which specify whether a refrigerator has one or two doors, as well as the presence of additional functionalities, such as "Minibar", "Usi_reversibile" and "Dozator_apa". The colour and finish of the product are captured through binary variables indicating whether the refrigerator is "Alb" (White), "Argintiu" (Silver), "Inox" (Stainless Steel), "Dark_Inox" (Dark Stainless Steel) or has a "Metal_Look" appearance. There are also fields dedicated to the manufacturer: "Producator" (string) specifies the general name, and the dataset includes a series of binary indicators for specific manufacturers (e.g., "Producator_Arctic", "Producator_Beko", "Producator_Heinner", "Producator_Tesla", etc.), thus facilitating the segmentation of products by brands. Regarding energy efficiency data, in addition to binary indicators for energy classes (e.g., "Clasa_Energetica_A", "Clasa_Energetica_E", "Clasa_Energetica_F", etc.), there is also a "Clasa_energetica" (string) field that consolidates the information in the form of a letter (for example, "E" or "F"), thus facilitating the classification of products according to energy consumption and environmental performance.

To ensure the quality and relevance of the data used in the analysis, several processing techniques were applied. First, missing values were treated by replacing missing numerical data with zero, and missing ratings were replaced with the average of available ratings to reduce their impact on prediction models. To capture more complex relationships between variables, derived variables were created, such as the volume/height ratio, squared terms for volume and height, and a variable measuring volume per door. These transformations were applied to improve the performance of the machine learning models used in the subsequent analysis. An essential step in data preparation was the detection and removal of extreme values, a process carried out using the Interquartile Range (IQR) method. In this regard, observations with prices exceeding the threshold established by IQR were eliminated, thus ensuring that the data used in the analysis is representative and not distorted by extremes. Also, given the distribution of refrigerator prices, a logarithmic transformation was applied to the price variable in order to reduce the asymmetry of the distribution and improve the performance of regression models.

In order to achieve our objective, the methodological framework employed a machine learning approach, utilizing an 80-20 stratified train-test split via the createDataPartition function to ensure representative sampling across the dataset. Feature normalization was implemented through Z-score standardization ($\mu=0$, $\sigma=1$) using the preProcess function with "center" and "scale" parameters, thereby establishing commensurable measurement scales across all variables and mitigating potential estimation bias.

To optimize models and reduce computational complexity, we implemented Label Encoding techniques and dimensionality reduction. For categorical variables, we applied label encoding by replacing multiple dummy variables with a single encoded numerical variable. For example, for manufacturers, instead of 10-15 separate variables (Producator_Arctic, Producator_Beko, etc.), we created a single numerical variable Producator_encoded with integer values (1, 2, 3, etc.). Similarly, we transformed the energy class from a categorical value (A+++, A++, etc.) into an ordinal numerical value, thus facilitating interpretation by machine learning models. For other categorical variables such as colour, refrigerator type, and control type, we applied the same transformation approach to numerical variants. For dimensionality reduction, we implemented multiple feature selection techniques. We used Spearman correlation to identify the most relevant features for price prediction, keeping only a subset of a maximum of 20 features with the highest absolute correlation. We eliminated zero-variance features, adding small perturbations in cases where elimination was not possible. In the selection process, we prioritized encoded features over one-hot representations, thus reducing redundancy. Additionally, we limited the derivation of new features to only two essential variables, the volume/height ratio and base area, instead of over 10 derived variables from previous iterations. As a result of these optimizations, we reduced the dataset dimensionality from 40-50 initial features to only 5-10 relevant features, including the main physical characteristics (volume, height), encoded variables for manufacturers and energy classes, the volume/height ratio, and variables with the highest correlation with price. This substantial dimensionality reduction of approximately 80% allowed for much faster model training and improved their performance by reducing noise and the risk of overfitting, thus contributing to an increase in the coefficient of determination R^2.

The analytical framework encompassed eight distinct predictive models. The initial models included Univariate Linear Regression utilizing available numerical predictors in isolation and Multivariate Linear Regression incorporating the comprehensive feature set. More complex algorithms were subsequently implemented, including Random Forest with optimized mtry hyperparameters governing feature subset selection and XGBoost Gradient Boosting Framework subjected to hyperparameter optimization across nrounds, max_depth, and eta (learning rate) parameters. Artificial Neural Networks with architectural variants spanning 5-15 hidden nodes were also evaluated, alongside Support Vector Machine Regression implementing radial basis function kernels with tuned sigma and C regularization parameters. The examination of regularization techniques involved Lasso Regression featuring L1 regularization with optimized lambda penalization coefficients and Elastic Net Regularization utilizing both L1 and L2 penalties with tuned alpha and lambda parameters.

Model optimization procedures implemented 10-fold cross-validation via the trainControl function, facilitating exhaustive hyperparameter tuning across the model spectrum. The Boruta algorithm was deployed for feature selection, identifying statistically significant predictors, including volumetric measurements, height parameters, energy efficiency metrics, and derived composite features such as volume-to-height ratios.

Performance evaluation incorporated multiple error metrics computed after inverse logarithmic transformation. Mean Absolute Error (MAE) was utilized for measuring absolute prediction deviation, while Root Mean Square Error (RMSE) quantified prediction variance with heightened sensitivity to large errors. Mean Absolute Percentage Error (MAPE) assessed relative prediction accuracy, and the Coefficient of Determination (R^2) evaluated explained variance proportion.

The analytical framework was augmented with several methodological refinements. Outlier identification and treatment utilized the Interquartile Range (IQR) methodology, while the elimination of zero-variance predictors enhanced model parsimony. Engineering of derived features, including volume-to-height ratios and volumetric door density, provided additional predictive capacity. Enhanced visualization of model performance through comparative graphical representations facilitated interpretation and hierarchical cluster analysis of model performance metrics and identified performance patterns across algorithmic approaches. Results were systematically documented and exported in multiple formats (Excel and CSV), providing comprehensive performance metrics, predictive outputs, and feature importance rankings to facilitate subsequent analysis and interpretation.

Results and discussions

The comparative analysis of predictive models for refrigerator pricing yielded significant findings regarding algorithmic performance differentials. XGBoost demonstrated superior predictive capability with the highest coefficient of determination ($R^2 = 0.7418$) and lowest error metrics (MAE = 472.1679 RON, RMSE = 734.9483 RON, MAPE = 22.3299%), followed closely by Random Forest ($R^2 = 0.7221$, MAE = 514.0647 RON, RMSE = 762.4469 RON, MAPE = 26.2206%). Support Vector Machine and Neural Network models exhibited moderate performance ($R^2 = 0.6772$ and $R^2 = 0.5116$, respectively), while regularized linear models (Elastic Net: $R^2 = 0.5235$; Lasso: $R^2 = 0.5488$) marginally outperformed standard Multiple Regression ($R^2 = 0.5516$). Simple Regression demonstrated substantially inferior performance ($R^2 = -0.0314$), indicating its inadequacy for this prediction task. The F-test for Multiple Regression revealed statistically significant relationships between predictor variables and refrigerator prices ($p < 0.01$).

The Boruta feature selection algorithm identified several statistically significant predictors influencing refrigerator pricing, with volume-related characteristics, brand indicators (particularly

Samsung, LG, and Bosch), construction type, premium features, and exterior material emerging as the most influential variables.

Statistical tests revealed significant violations of classical regression assumptions, with the Shapiro-Wilk test indicating non-normal distribution of residuals ($p < 0.01$), the Breusch-Pagan test confirming heteroscedasticity ($p < 0.01$), and the Durbin-Watson test indicating autocorrelation in the residuals ($p < 0.05$).

Analysis of percentage errors across different price segments revealed a notable pattern, with XGBoost demonstrating a median percentage error of approximately 22.3299% for refrigerators priced below 1000 RON, progressively decreasing to approximately 15% for the premium segment (>3000 RON). The k-means clustering approach applied to model performance metrics identified distinct performance tiers, with XGBoost and Random Forest forming the highest-performing cluster, as evident in Table 1.

Table 1. The performance of each model used in our analysis

Model	MAE (RON)	MSE (RON²)	RMSE (RON)	MAPE (%)	R2	Cluster
XGBoost	472.1679	540,148.935	734.9483	22.3299	0.7418	2
Random Forest	514.0647	581,325.2715	762.4469	26.2206	0.7221	2
SVM	561.2853	675,226.0892	821.7214	28.7494	0.6772	2
Multiple Regression	670.1426	938,046.965	968.5282	36.7663	0.5516	3
Lasso	679.8348	943,829.6172	971.5089	38.0319	0.5488	3
Elastic Net	692.3946	996,760.3881	998.3789	38.8122	0.5235	3
Neural Network	688.024	1,021,718.942	1,010.8011	34.2975	0.5116	3
Simple Regression	1,095.8279	2,157,587.3094	1,468.8728	62.8353	-0.0314	1

Source: Authors' own research contribution.

The empirical findings of this study provide significant insights into the price determination mechanisms within the refrigerator market. The superior performance of tree-based ensemble methods, particularly XGBoost ($R^2 = 0.7418$) and Random Forest ($R^2 = 0.7221$), compared to traditional econometric models, confirms the presence of complex non-linear relationships in price formation that cannot be adequately captured through linear specifications, as illustrated in Figure 1. This finding aligns with Xing et al. (2004), who demonstrated similar algorithmic superiority in consumer electronics pricing models and extends Mitra's (2022) work, which emphasized the limitations of parametric approaches in capturing multidimensional product differentiation effects.

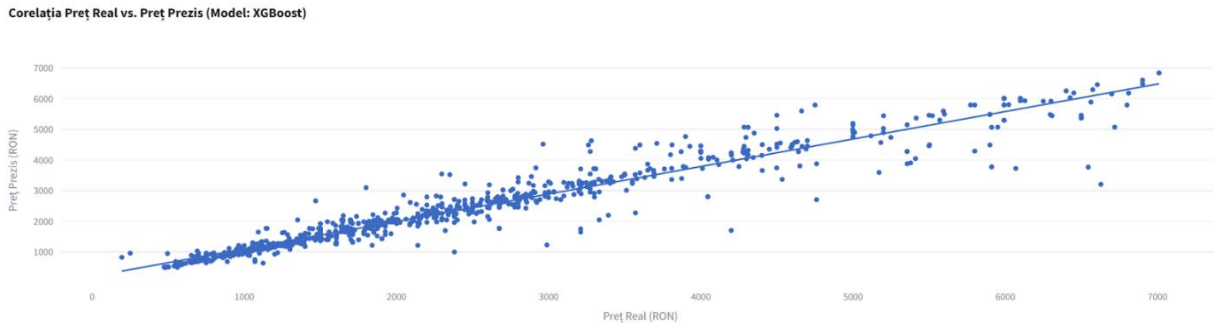

Figure 1. Correlation between the actual price and the predicted price

Source: Authors' own research results.

The substantial violations of classical regression assumptions—heteroscedasticity, non-normality of residuals, and autocorrelation—further support the economic intuition that pricing structures in differentiated product markets follow complex patterns reflecting strategic positioning decisions rather than simple cost-plus formulations.

Feature importance analysis offers valuable economic insights regarding the market valuation of product attributes, as can be inferred from the correlation heatmap in Figure 2. The significant influence of brand variables, particularly premium manufacturers, quantitatively confirms Xiao and Li's (2023) theoretical framework on brand equity premiums in durable goods markets. The economic interpretation suggests that consumers perceive substantial quality differentials between manufacturers, with established premium brands commanding significant price premiums independent of objective technical specifications, as evidenced in the boxplot distributions shown in Figure 3. This conclusion contradicts Istudor's (2023) finding that brand effects diminish when controlling comprehensively for product characteristics, suggesting instead that brand equity represents genuine economic value beyond measurable attributes.

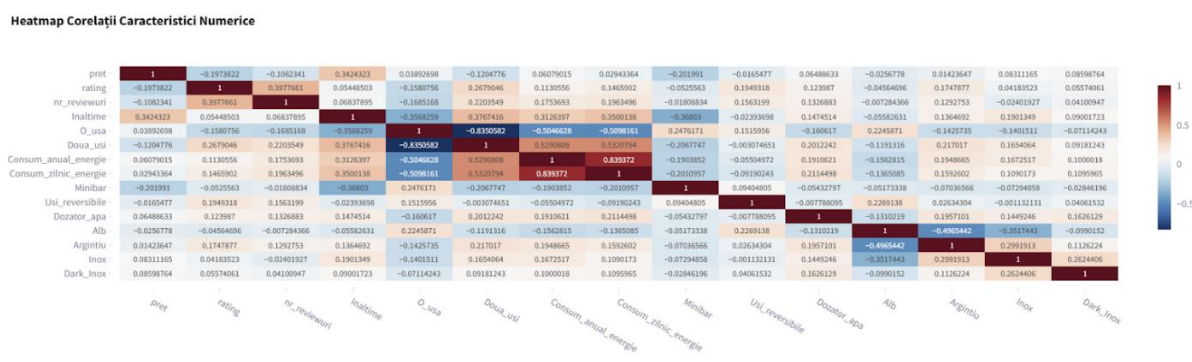

Figure 2. The correlation heatmap of the variables

Source: Authors' own research.

DOI: 10.2478/picbe-2025-0203, pp. 2639-2649, ISSN 2558-9652 |
Proceedings of the 19th International Conference on Business Excellence 2025

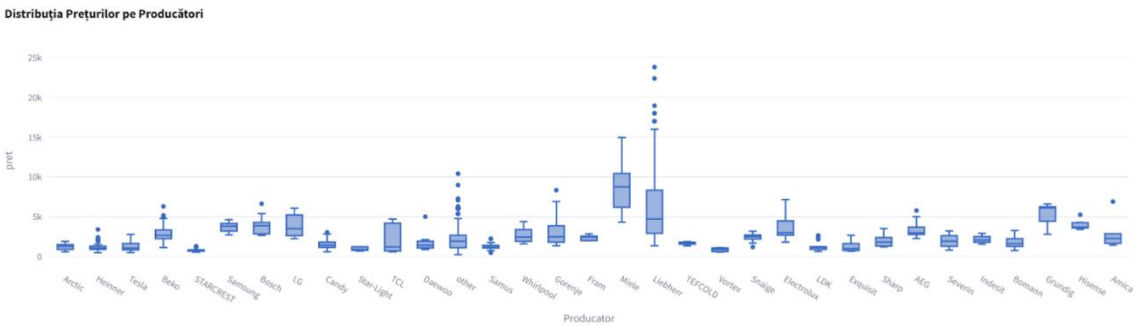

Figure 3. The boxplot for the price distribution for each producer

Source: Authors' own research results.

The observation that prediction errors systematically vary across price segments provides empirical evidence supporting market segmentation theories, as demonstrated in Figure 4. The higher prediction accuracy in premium segments (15% error) compared to budget segments (22.3299% error) indicates more consistent pricing mechanisms in upmarket positions, supporting Chen et al.'s (2021) observation that lower-priced segments exhibit greater price competition and promotional variability. This pattern suggests that manufacturers employ different economic strategies across market segments—value-based pricing dominates the premium segment, while cost-plus and competitive pricing strategies prevail in budget categories.

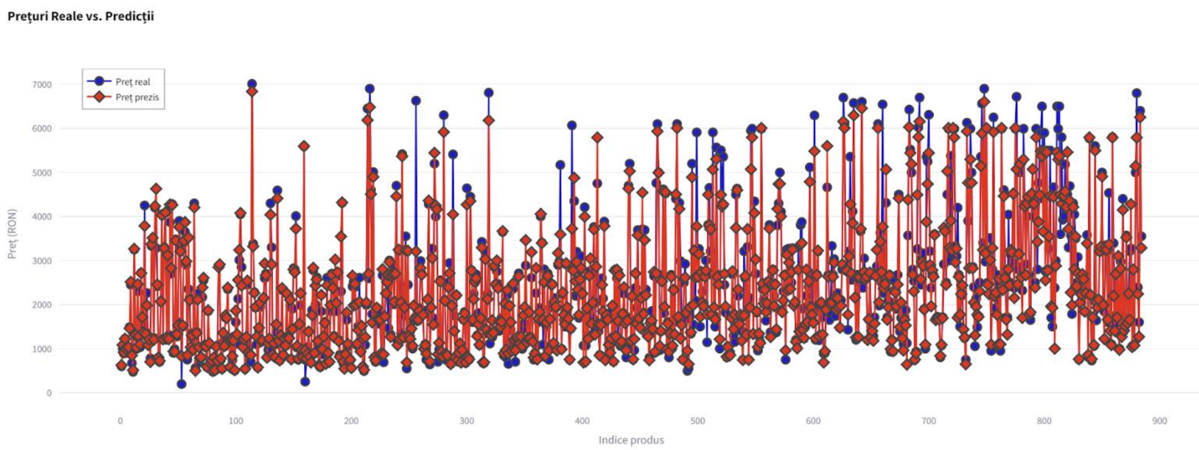

Figure 4. The difference between the real prices (blue) and the predictions (red)

Source: Authors' own research results.

The clustering approach to model evaluation demonstrates that algorithmic performance follows distinct patterns across metrics, providing methodological validation for Mitra's (2022) argument that economic model selection should incorporate multiple performance dimensions simultaneously. The moderate improvement from regularization techniques (Elastic Net $R^2 = 0.5235$ vs Multiple Regression $R^2 = 0.5516$) suggests that multicollinearity issues exist but are not the primary limitation in linear specifications, contradicting Xing et al. (2004) emphasis on collinearity as the dominant concern in hedonic price models. Instead, our findings support Istudor et al. (2023) argument that functional form misspecification represents the more fundamental limitation in parametric approaches to differentiated product markets.

DOI: 10.2478/picbe-2025-0203, pp. 2639-2649, ISSN 2558-9652 |
Proceedings of the 19th International Conference on Business Excellence 2025

This research extends previous economic modelling approaches by quantifying the specific non-linear relationships between product attributes and market prices, demonstrating that machine learning algorithms can effectively capture the complex equilibrium resulting from strategic pricing decisions in oligopolistic markets characterized by multidimensional product differentiation. The comparative performance of all models in terms of R^2 is visualized in Figure 5, which clearly illustrates the superiority of ensemble methods (XGBoost and Random Forest) over traditional approaches, with Simple Regression showing the poorest predictive power.

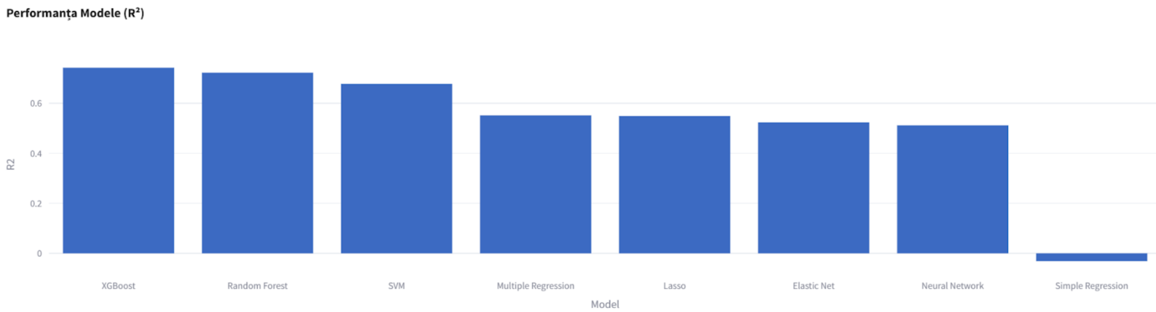

Figure 5. The value of the R^2 for each performance model

Source: Authors' own research results.

Conclusion

The predicted price provides valuable insights for sellers and resellers (such as eMAG) regarding the perceived value of a product based on its characteristics. This information enables them to adjust the selling price accordingly. If the original price is lower than the predicted value, the seller may increase the price without significantly losing market share, as the product's features justify a higher selling price. Conversely, if the original price is higher than the predicted value, a price reduction may be necessary to stimulate sales and gain market share, as the product is likely overvalued and requires a price adjustment.

While adding more product characteristics could enhance the model's robustness, it is not strictly necessary. The selected features represent common attributes found across refrigerators listed on eMAG. However, inconsistencies in product specifications limited the number of shared characteristics that could be analyzed across all 964 refrigerators, preventing a more extensive feature set. Despite these constraints, the model remains reliable due to the inclusion of key product attributes.

Sellers and resellers require an estimated product value to set an optimal selling price that ensures both market competitiveness and long-term profitability. Our system provides data-driven price predictions based on product characteristics and statistical analysis, enabling informed decision-making. However, it does not account for external business factors such as competition, brand positioning, or marketing strategies, all of which can significantly influence pricing. As a result, while our predictions assist in price adjustments, they do not prescribe exact increases or decreases in selling prices.

Despite these insights, the study presents certain limitations that must be taken into consideration when interpreting the results. The developed models depend on the quality and completeness of the extracted data, while market dynamics and fluctuations can reduce the accuracy of long-term predictions. For future research, we propose expanding the study in multiple directions. Implementing time series analysis methods such as LSTM and ARIMA would allow us to capture and model seasonal price fluctuations. Exploring advanced neural network architectures could as well improve prediction accuracy for products with complex characteristics. Additionally, integrating longitudinal data would help capture seasonal price variations, while combining data from various e-commerce platforms would reduce bias specific to a single source. The implementation of deep learning algorithms would be particularly beneficial for improving prediction accuracy for products with complex feature sets.

Acknowledgements

The research study has been elaborated within the Data Science Research Lab for Business and Economics of the Bucharest University of Economic Studies. This paper was co-financed by the Bucharest University of Economic Studies during the PhD program. This work was supported by a grant from the Bucharest University of Economic Studies, through the project 'Analysis of the Economic.

References

Asghar, M., Mehmood, K., Yasin, S., & Khan, Z. M. (2021). Used cars price prediction using machine learning with optimal features. Pakistan Journal of Engineering and Technology, 4(2), 113-119.

Deng, S., Zhu, Y., Huang, X., Duan, S., & Fu, Z. (2022). High-frequency direction forecasting of the futures market using a machine-learning-based method. Future Internet, 14(6), 180.

Ho, W. K., Tang, B. S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. Journal of Property Research, 38(1), 48-70.

Istudor, N., Dumitru, I., Filip, A., Stancu, A., Roșca, M. I., & Cânda, A. (2023). Integration of circular economy principles in consumer behaviour for electrical and electronic equipment. Amfiteatru Economic, 25(62), 48-62.

Li, H., Xiao, Q., & Peng, T. (2023). Optimal pricing strategy of new products and remanufactured products considering consumers' switching purchase behavior. Sustainability, 15(6), 5246.

M'hamdi, O., Takács, S., Palotás, G., Ilahy, R., Helyes, L., & Pék, Z. (2024). A comparative analysis of XGBoost and neural network models for predicting some tomato fruit quality traits from environmental and meteorological data. Plants, 13(5), 746.

Mitra, S. (2022). Economic models of price competition between traditional and online retailing under showrooming. Decision, 49(1), 29-63.

Orji, U., & Ukwandu, E. (2024). Machine learning for an explainable cost prediction of medical insurance. Machine Learning with Applications, 15, 100516.

Xing, X., Tang, F. F., & Yang, Z. (2004). Pricing dynamics in the online consumer electronics market. Journal of Product & Brand Management, 13(6), 429-441.

Zehtab-Salmasi, A., Feizi-Derakhshi, A.-R., Nikzad-Khasmakhi, N., Asgari-Chenaghlu, M., & Nabipour, S. (2020). Multimodal price prediction.

Safonov, K. (2024). Neural network approach to demand estimation and dynamic pricing in retail.

Hasnain, M., Sajid, A., & Awan, M. A. (2024). Predicting the price of used electronic devices using machine learning techniques.