urban science

MDPI

Review

Comparative Analysis of Advanced Models for Predicting Housing Prices: A Review

Inmaculada Moreno-Foronda [1], María-Teresa Sánchez-Martínez [1] and Montserrat Pareja-Eastaway [2,*]

[1] Department of Applied Economics, University of Granada, 18011 Granada, Spain; inmamf@correo.ugr.es (I.M.-F.); tsanchez@ugr.es (M.-T.S.-M.)
[2] Department of Economics, University of Barcelona, 08034 Barcelona, Spain
* Correspondence: mpareja@ub.edu

Abstract: Understanding the determinants of housing price movements is an ongoing subject of debate. Estimating these determinants becomes a valuable tool for predicting price trends and mitigating the risks of market volatility. This article presents a systematic review analyzing studies that compare various machine learning (ML) tools with hedonic regression, aiming to assess whether real estate price predictions based on mathematical techniques and artificial intelligence enhance the accuracy of hedonic price models used for valuing residential properties. ML models (neural networks, decision trees, random forests, among others) provide high predictive capacity and greater explanatory power due to the better fit of their statistical measures. However, hedonic regression models, while less precise, are more robust, as they can identify the housing attributes that most influence price levels. These attributes include the property's location, its internal features, and the distance from the property to city centers.

Keywords: machine learning; hedonic prices; prediction; prices; housing

check for updates

Academic Editor: Maria Rosaria Guarini

Received: 23 December 2024
Revised: 23 January 2025
Accepted: 29 January 2025
Published: 31 January 2025

Citation: Moreno-Foronda, I.; Sánchez-Martínez, M.-T.; Pareja-Eastaway, M. Comparative Analysis of Advanced Models for Predicting Housing Prices: A Review. *Urban Sci.* **2025**, *9*, 32. https://doi.org/10.3390/urbansci9020032

1. Introduction

Real estate assets serve a dual purpose: they are an essential consumer good, and their acquisition represents the most significant investment in the portfolios of most households. Therefore, a thorough understanding of housing price predictability is essential for making informed investment decisions and ensuring the efficient allocation of assets.

Predicting housing prices is crucial not only for households and other market participants but also for policy design, economic research, and the broader economy, given the macroeconomic significance of residential price trends [1]. Sustained increases in housing prices pose a serious problem in virtually all countries, as they make it difficult for many citizens to access housing.

However, predicting housing prices is complex due to the significant challenges in measuring all the intrinsic and extrinsic characteristics that form part of a property [2]. Among the different variables/factors that influence housing prices are, on the one hand, the internal variables or micro factors, which refer to the structural characteristics of the property itself. These include specific features of the property, such as size, number of rooms, or type of construction, as well as specific characteristics of the building and common areas that are part of it (elevator, natural gas installations, swimming pool, or park, among others). On the other hand, there are external variables or macro factors that encompass all types of location and neighborhood attributes of the property, among which accessibility and distance to essential services, the economic and social conditions of the district where it is located, or the particularities of urbanization of the surroundings, such as sidewalks, lighting, etc., stand out [3]. Finally, and no less important, are the regional factors, which refer to the exact location where the property is located, including transportation conditions, public support facilities, commercial prosperity, etc., as well as the socio-economic characteristics associated with the geographic location, although the latter are difficult to quantify [4].

There are two general approaches to evaluating prices in the real estate market. The first is the traditional method, where the value of a property is estimated through the appraisal or opinion of a trained and experienced expert in the field. This method employs comparison, cost, accounting, development, and capitalization methods, among others, which inevitably retain a high degree of subjectivity [5].

The second approach involves advanced prediction models, such as the hedonic price model (HPM), machine learning models, case-based reasoning, and spatial analysis methods.

Among advanced price prediction models, hedonic price models (HPMs) stand out. HPMs have been more widely adopted by academics and professionals for property valuation in various real estate markets worldwide. These models measure the contribution of property attributes as well as other external factors that could influence the value of a property [6].

However, over time, and with the continuous improvement of technologies and mathematical and statistical procedures, new advanced property valuation methods have been developed, offering results that are closer to reality. The use of artificial intelligence enables the creation of fully objective prediction models, such as artificial neural network models, random forest models, or support vector machines, among others [7].

In recent decades, advancements in technology and statistical procedures have enabled the acquisition and storage of larger volumes of data. Big data applications are attracting growing interest among urban researchers. Notable developments include advances in data engineering and the use of machine learning techniques for applications across nearly all economic sectors. Machine learning is a branch of artificial intelligence, emerged with the aim of addressing the issue of subjectivity when determining housing prices [8]. It is one of the most widely used methods for predicting housing price models, through which computers identify patterns in data without being explicitly programmed for that task [9]. This methodology imitates human reasoning, and due to its problem-solving capabilities, the models it encompasses are efficient alternatives for housing price prediction, as they can identify non-linear relationships among relevant data. The development of these models can significantly aid in forecasting future housing prices and in establishing real estate policies [10,11].

The use of this tool, which does not rely on socioeconomic behavior models, raises doubts about its effectiveness in providing better predictions than traditional methodologies [12].

Despite growing concerns about the use of new prediction models in the real estate sector, there are no systematic studies identifying whether these pricing measurement methodologies are actively used in the real estate sector. Nor is there a review of which models are most widely used, which provide the most accurate results in predicting housing prices, in which countries these studies are conducted, or which of them are carried out with real data and for experimental purposes. Furthermore, despite the increasing use of machine learning models for housing price prediction, there are no studies that compare several predictive models for real estate in a specific geographic area.

The general aim of this paper is to address this gap in the literature, along with a series of specific objectives: First, a systematic review is conducted to understand the different advanced models used for housing price prediction and to determine in which countries these models are most frequently used, what methods experts employ, and which models provide the most significant results when predicting real estate prices. Second, the paper seeks to compare several advanced prediction models and their consideration of the different internal and external characteristics that influence the establishment of a property's value. Finally, the paper aims to present the results of comparing different methods applied to real cases in the housing market across different international regions for housing suppliers and institutional entities when formulating their economic policies.

The structure of the article is as follows: the next section provides an overview of the literature on advanced real estate valuation and prediction models, followed by a description of the methodology employed for the systematic review of the identified articles. Lastly, an analysis and discussion of the results are presented.

2. Advanced Housing Price Prediction Methods: A Classification

This section presents a classification of different advanced approaches for real estate price valuation: HPMs, neural networks, spatial analysis, etc.

Hedonic pricing models (HPMs) date back to 1939 when Andrew Court succeeded in estimating an index for the automotive market using this model [13]. Technically, the theory behind HPMs originates from Lancaster's renowned consumer theory and Rosen's theoretical model [14]. Rosen stood out for being the first to apply HPMs within standard economic theory, incorporating properties considered heterogeneous in nature. Over the past decades, this model has been widely used in housing market research due to the high heterogeneity of housing characteristics [15].

HPMs aim to establish, through implicit price equations, the value of a property based on its attributes including structural attributes (quality, age, and size of the building), location attributes (proximity to transportation links, employment centers, and services), and neighborhood attributes (crime levels and socioeconomic profile) [15]. This approach explains its relevance in real estate market studies by considering housing as a heterogeneous good composed of various features. The hedonic methodology is considered a viable way to study the real estate market, allowing for the estimation of the values of each housing attribute that buyers and sellers consider during the transaction process—an aspect that can be highly useful to stakeholders [16].

Artificial intelligence (AI) and its mathematical models have also been applied to new housing price valuation and prediction models. Machine learning, a branch of AI, has been under development since the 1950s. In 1980, a breakthrough occurred with the Hopfield model, the first neural network created by John Hopfield, whose neurons were all input and output, fully connected to each other in both directions [17].

In the 1980s, Hinton and Sejnowski developed machines that combined Hopfield neural networks with the simulated annealing algorithm, resulting in a two-layer fully connected neural network [18]. Around 1995, Vapnik and Cortés achieved one of machine learning's most significant advancements: Support Vector Machines (SVM). These machines map input vectors non-linearly into a high-dimensional feature space, enabling the construction of a linear decision boundary and yielding accurate results [19].

At the beginning of the 21st century, Breiman (2001) introduced another machine learning method, Random Forest, which combines several decision trees. Each tree is created from a random subset of features, with nodes generated using randomly selected characteristics [20].

Machine learning models have advanced further, exploring new directions. Combining multiple models has led to new methodologies widely used by researchers across various fields, including housing price prediction. In these models, computers identify patterns in data without being explicitly programmed for the task, mimicking human reasoning [9]. The most commonly used models include linear and logistic regression, decision trees, Random Forest, Bayesian networks, AdaBoost, Support Vector Machines, XGBoost, K-Nearest Neighbors, and Artificial Neural Networks [21].

In Figure 1, a classification of the different advanced approaches for valuing real estate prices is presented: HPM, neural networks, spatial analysis, and so on.

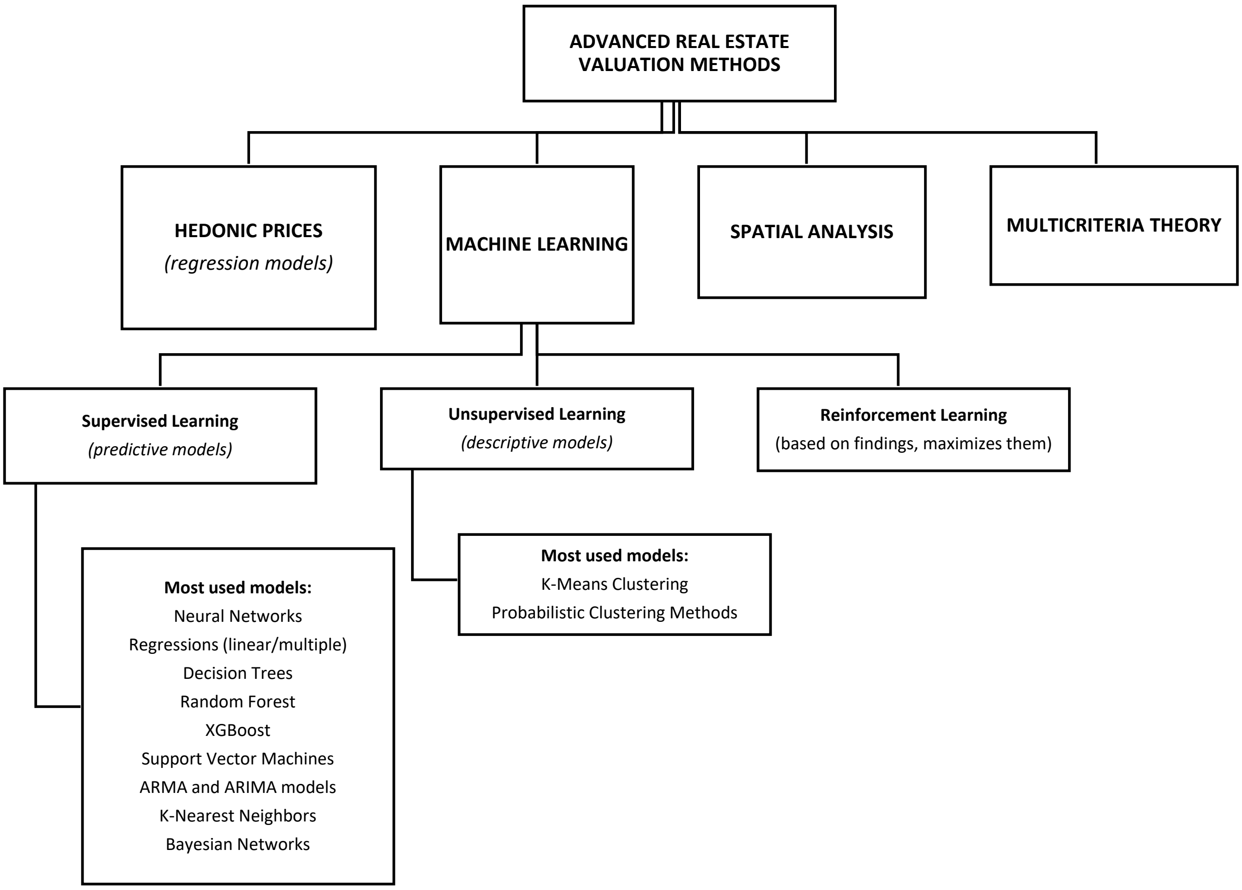

Figure 1. Most commonly used advanced prediction models (source: own elaboration).

Advanced valuation methods in the real estate sector are classified into four categories: hedonic price models (regression models that estimate the value of each characteristic of a property), machine learning models (based on mathematical techniques and artificial intelligence), spatial analysis models (which focus on estimating the characteristics of a specific area), and multi-criteria theory models (which estimate all types of intangible property characteristics).

Machine learning is classified into three types of learning: supervised learning, which consists of predictive models and forms the basis of the analysis in this review (neural networks, decision trees, XGBoost, etc.); unsupervised learning, which focuses on analyzing and predicting descriptive models; and reinforcement learning, which aims to maximize findings.

Thanks to their problem-solving capabilities, the models that make up these methods are efficient alternatives for housing price prediction, as they can identify non-linear relationships between relevant data. The development of these models helps predict future housing prices and establish real estate policies [10].

3. Methodology for the Systematic Review of Housing Price Estimation Models

A systematic review integrates various studies that analyze the same question, considering these studies in an observational manner to assess the trajectory of the analysis on the topic being reviewed. The main objective of a systematic review is to gather the highest number of original studies on the topic being analyzed, in order to provide a high level of scientific evidence [22].

Once the relevant studies for the research are located, they are analyzed comprehensively to compare them and identify their most important aspects [23]. Systematic reviews provide a fusion of information on a specific topic, which is essential for future research on the same subject and can offer answers to questions that could not be obtained through independent studies. Thus, systematic reviews are highly useful to researchers when addressing the situation of a topic they wish to study and analyze [24,25].

Systematic reviews are often conducted with the help of the PRISMA Statement. It is essentially a guide to assist authors in conducting this type of study [26].

This study used a systematic review to examine the number of existing articles that present practical cases of various advanced prediction models, specifically in the real estate sector, in a comparative manner, seeking to identify which of these models offers the best results for predicting housing prices (see Figure 2).

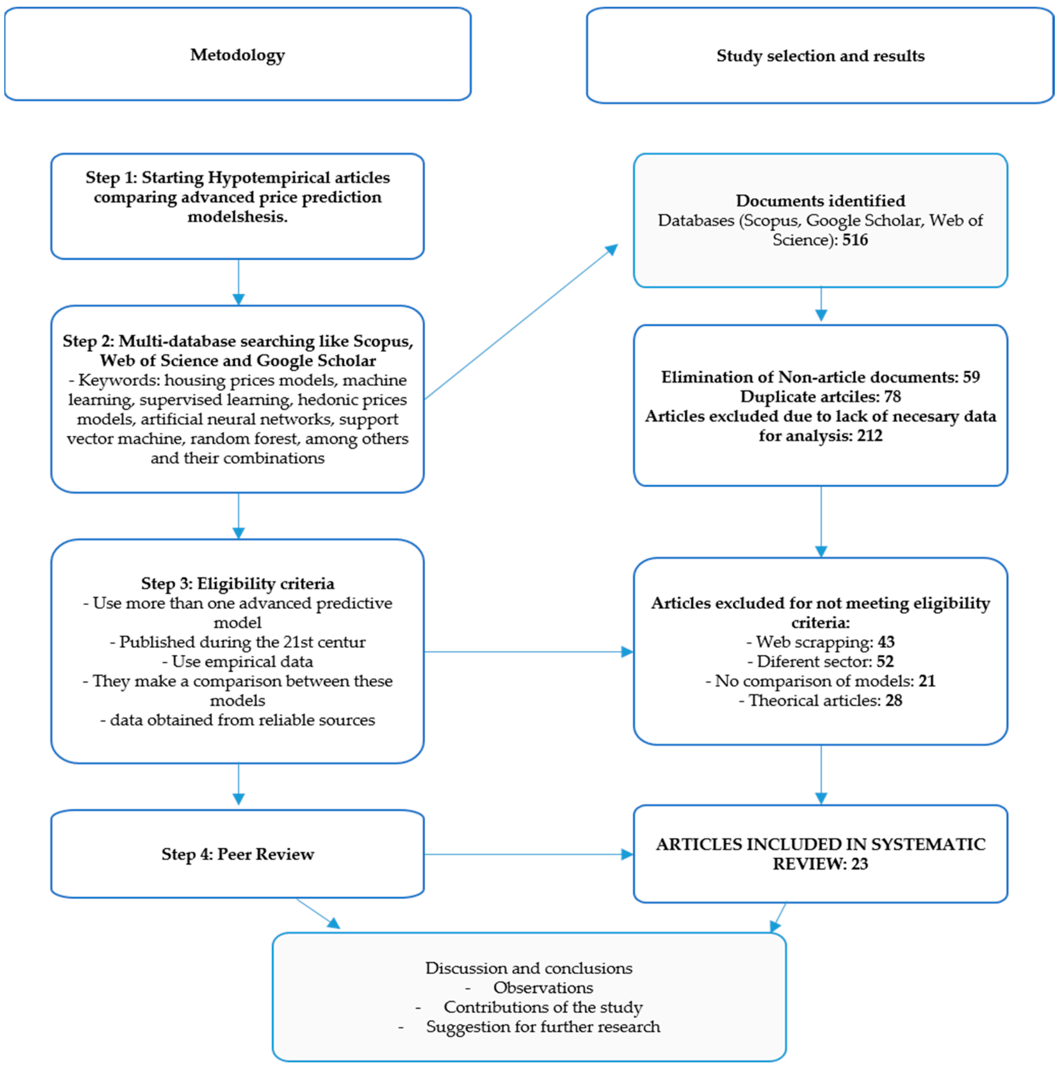

Figure 2. Flowchart of the article selection process for systematic review—PRISMA (source: own elaboration).

The articles used for this review were retrieved electronically from the Scopus, Google Scholar, and Web of Science platforms in May 2024; more specifically, the search was conducted in Scopus from 1 May to 10 May, in Google Scholar from 10 May to 20 May, and in Web of Science from 20 May to 30 May. These three platforms are important databases for retrieving studies, as they hold a wide range of documents categorized across different scientific fields.

Several combinations of keywords in English were entered into these three databases to locate as many studies on the topic as possible for analysis in the review. The terms used, among others, for the search were "housing price prediction"; "housing price models"; "machine learning"; "housing prices"; "supervised learning", "hedonic prices models", "artificial neural networks", "support vector machine", and "random forest".

3.1. Search Criteria

The exploration of the literature on the topic of interest was carried out longitudinally over time, searching for studies that have been published since the early 21st century, including preprints. The search was conducted so that the keywords in the documents would appear in the title, abstract, or in the keywords of the articles. Similarly, only documents in English or Spanish were selected. Only articles were chosen, excluding book chapters, technical reports, dissertations, editorials, opinions, and conference proceedings.

3.2. Eligibility Criteria

The articles included in the review meet the following requirements:

- They conduct housing price prediction studies using more than one advanced predictive model.
- They have been published in the 21st century, i.e., from the year 2000 to 2024, to track advancements in the use of these techniques over time.
- They use empirical data, meaning the models have been applied practically and are not just explained theoretically.
- They provide a comparison between these models.
- The data used must be obtained from reliable sources, whether public organizations or companies dedicated to the real estate sector, to provide empirical housing price data.

Articles that use data obtained through the "web scraping" technique are excluded, as this method provides potential sale price data rather than the actual transaction value of the property.

4. Study Selection and Results

The search for documents on housing price prediction models yielded a total of 516 studies. After conducting the necessary screening, excluding duplicate articles and those that did not meet the inclusion criteria, a total of 23 articles were selected for inclusion in this systematic review. The selection of articles has been supervised by two reviewers throughout the procedure, including a third reviewer in the final phase to resolve any disagreements that the first two might have.

Figure 2 shows the process of selecting the articles included in this review and highlights those that were excluded. After removing non-article studies and duplicate articles, further exclusions were made for various reasons. First, 212 articles were eliminated for lacking essential requirements for the analysis, such as failure to identify the source of the data or the year in which the study was conducted, among others. The remaining records were then subjected to eligibility criteria, resulting in the exclusion of 43 articles for being conducted using the "Web Scraping" technique; 52 articles were excluded because they focused on prediction models for fields other than real estate, such as mobile telephony or tourism, among others; 21 articles were excluded for not comparing two or more prediction models; and finally, 28 articles were removed for presenting only the theoretical part of the models without conducting an experimental analysis. In this way, the 23 articles included in this systematic review were obtained.

4.1. General Characteristics of the Included Studies

Table 1 outlines the main characteristics of the articles included in this systematic review. In terms of the number of authors, it is notable that 31% of the selected articles were co-authored by three authors, and 26% by four authors. Despite the challenges involved in conducting predictive price analysis for housing—due to the complexity of obtaining, tabulating, and adapting data, as well as executing predictive models in computer applications—it is worth highlighting that 22% of the articles were written by a single author. A smaller percentage, 17%, had two co-authors, and only one article was authored by six people, representing 4% of the total articles included in the review.

Table 1. Characteristics of the articles selected for the review.

Number of Articles Included	**23**
Authors	
1 author	5
2 authors	4
3 authors	7
4 authors	6
5 authors	0
6 authors	1
Country	
Kosovo	1
South Korea	2
Australia	2
Poland	1
USA	3
Taiwan	4
United Kingdom	1
Turkey	2
China	2
Slovenia	1
Thailand	2
Spain	2
Year from publication	
From 2000 to 2010	1
From 2011 to 2020	8
From 2021 to 2024	14
Type of study	
Cross-sectional	8
Longitudinal	15
Data Sample Analysis Period	
From 1 month to 5 years	17
From 6 years onwards	6
Sample Size	
From 0 to 5000	9
From 5000 to 10,000	3
From 10,000 onwards	11

Table 1. *Cont.*

Number of Articles Included	**23**
Number of models compared in the study	
Two models	7
Three models	7
Four models	4
Five models	2
Seven models	1
Eight models	1
Eleven models	1

Source: own elaboration.

Most empirical applications have been carried out in Asia, comprising 43% (10 studies), with Taiwan being the country that has published the most research on advanced housing price prediction models. A total of 26% of the articles were from European countries, 13% from America, specifically the United States, and 9% from both Australia and Turkey (Middle East), which are the geographical regions with the fewest articles. As observed, the Asian continent is a pioneer in this type of methodology in the housing market due to its significant technological advancements and the ease of obtaining macro data from official organizations [27].

The study design was predominantly longitudinal (15 out of 23 articles), while only eight were cross-sectional. Regarding the time periods of the data used in the analysis, 74% of the articles collected data over periods ranging from one month to five years, while 26% focused on longer periods, starting from six years onward.

Concerning the comparative study analysis between advanced prediction models, seven articles compared the results of using two models, seven others compared three models, four articles examined four models, and two articles compared five models. Additionally, two articles compared seven and eight models, respectively, while one article compared up to eleven advanced prediction models.

Finally, it is worth noting that most of the publications were made from 2020 onwards, with 61% of the articles published during this period. This could be attributed to advancements in technologies and mathematical techniques in recent years, which have led to a surge in comparative studies of price prediction models, aiming to better understand how both internal and external variables influence housing price fluctuations. Only one article from the early 2000s was included in this systematic review, with the rest published between 2011 and 2020.

4.2. Advanced Housing Price Prediction Models

In the previous section, it is noted that a large number of comparative models have been used in the various articles, with some including up to seven or eight prediction models. In fact, one study compares up to eleven models, primarily due to the number of observations involved in the analysis of each method. On the other hand, many articles compare only two or three models.

In total, twenty-three different prediction models have been used to carry out variations in housing prices based on both internal and external characteristics, including, in some studies, variables related to the neighborhood, location, availability of surrounding services, and environmental characteristics. Technological advancements have enabled the use of applications that enhance the prediction of these models, given the large number of observations they include, allowing for research that is highly useful for entrepreneurs, investors, and financial institutions interested in understanding fluctuations in real estate prices based on specific characteristics. In fact, many of these are variations and combinations of the most widely used models, as shown in Table 1.

Figure 3 illustrates the types of models used in the studies of this review, as well as how often these models appeared in the various articles. The most frequently used model, appearing in fourteen different articles, is the Random Forest model. This model consists of the creation and combination of several decision trees, with the aim of combining the outputs of these trees into a single output to address regression or classification problems [28]. In fact, in five of the thirteen articles where it appears, this model was chosen as the one that provides the most significant results compared to the others, due to statistical results such as R^2 or the absolute error.

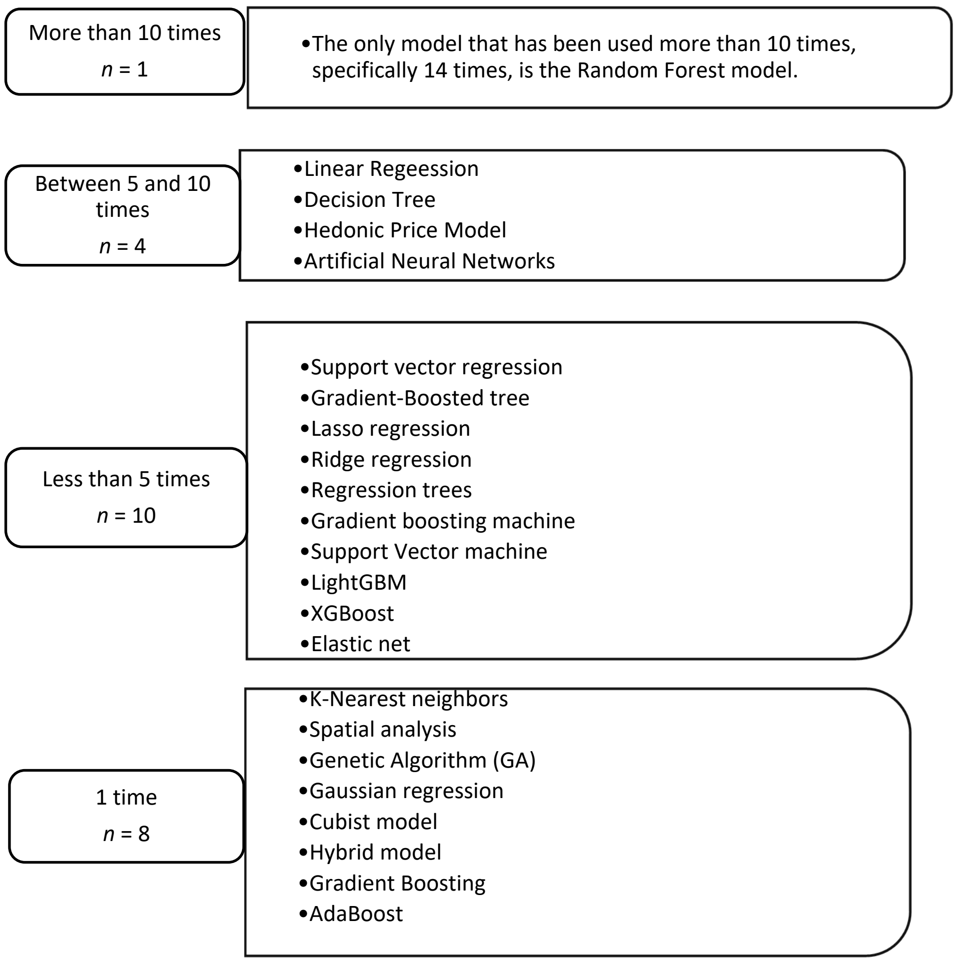

Figure 3. Number of times prediction models appear in the articles (source: own elaboration).

The models that appear most frequently (between five and ten times) in the different articles are the most commonly used in prediction studies. These include regression models, decision trees, HPMs, and artificial neural networks.

Next, there are ten models that appear in fewer than five articles, either because they are less well-known or because their predictions are not as accurate or significant as the previous ones. These include XGBoost, regression trees, support vector machines, and Lasso regression, among others. Finally, seven models are the least frequent and least known in the field of housing price prediction, as they appear in only one article.

4.3. Most Influential Variables in Housing Price Prediction Models

To use advanced housing price prediction models, it is necessary to include not only a sample of housing transactions but also a significant number of variables (both internal and external) related to the properties in order to understand how prices vary based on these aspects. Figure 4 shows the variables from the articles in the systematic review that have proven to be the most significant and, therefore, have had the greatest impact on the process of housing price variation and prediction.

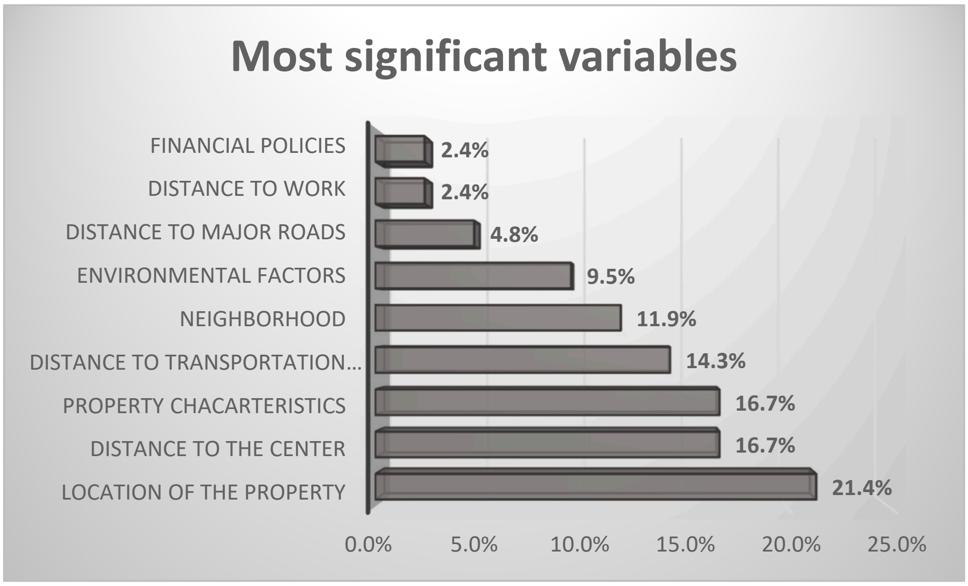

Figure 4. Most influential variables in housing price variation in the articles of the systematic review (source: own elaboration).

According to Figure 4, the variables that have proven to be the most significant and influential on housing prices, and those that make properties more attractive to buyers according to the models analyzed in the selected articles, are nine. The variable that has been repeated the most (up to nine times) and, therefore, can be considered the most influential, is the location of the property, which includes the area/neighborhood in which it is located, as well as the characteristics of that area.

Next, following the order of importance, there are two variables that have appeared seven times each: distance to the central commercial district, which includes the distance to the city center and the services it offers, such as shopping centers, hospitals, educational institutions, etc.; structural characteristics of the property, which involve factors such as the size, number of rooms, presence of an elevator, outdoor spaces, and others.

The variable considering distance to transportation infrastructure, i.e., the distance to train stations, trams, buses, etc., has been one of the most significant, appearing in six studies. In five of the articles, the neighborhood and its characteristics were considered one of the most influential variables in the models. This variable includes economic, ethnic/racial conditions, etc.

Environmental factors of the location, which include green spaces, gas emissions, and noise levels, among other aspects, in the geographical area where the property is located, have been significant in four articles. A factor that has been repeatedly mentioned twice as influential is the distance from the property to major roads. Lastly, with only one incidence, two variables appear: distance from the property to workplaces and financial policies, with the latter including the conditions set by banks when applying for a mortgage loan for property acquisition.

5. Discussion

In order to present all the studies selected for this systematic review, Appendix A provides a concise, detailed overview of these studies, including authors, the period in which the analysis took place, sample sizes, compared prediction models, and a brief summary/conclusions of the objective of each study.

In the selected articles for this systematic review, the various authors choose those models whose results provide the best fit for housing price prediction based on the corresponding parameters. They focus on statistical measures that demonstrate this fit, namely selecting the model that achieves both a higher R^2, which measures how well the model fits the data, as well as the lowest levels of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Integrating both MAE and RMSE into prediction models helps to better understand the ability of different models to predict prices more accurately [29].

In Figure 5, the different particularities, both positive and negative, of the main advanced price prediction models in the housing market are presented, such as the hedonic price models and machine learning models, as discussed throughout this article. These characteristics are analyzed through a SWOT analysis, commonly used to identify weaknesses, threats, strengths, and opportunities. This will help to understand the positive and negative aspects of the main advanced prediction methods and allow for a thorough comparison of them.

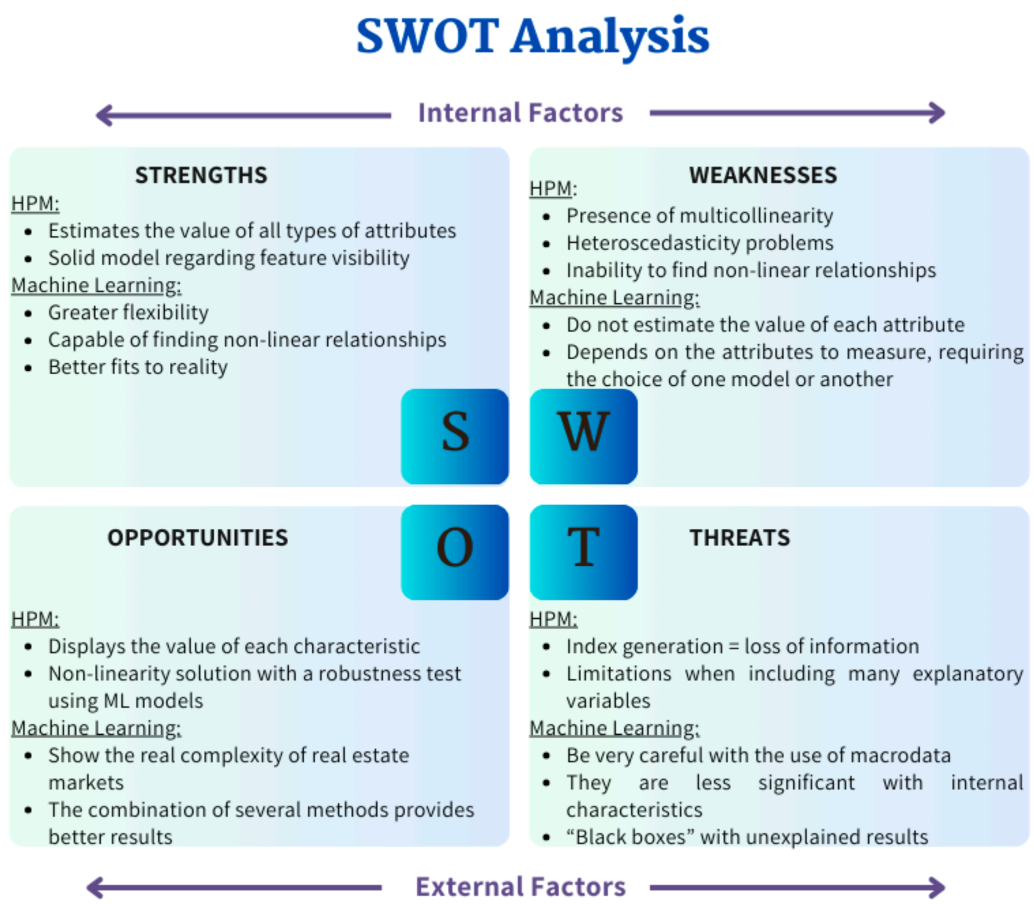

Figure 5. Characteristics of HPMs and machine learning models through the SWOT analysis (source: own elaboration).

In the works of [2,30–36], where all authors focus on the analysis of HPM, they agree that, when compared to machine learning models, the latter offer better adjustments in prediction models. According to these authors, this is due to the common occurrence of multicollinearity in HPM and the generation of indices that this model requires, which leads to a loss of information when interpreting results. In some cases, HPM can also face heteroscedasticity problems, though these can be corrected, and the variables may still be significant. HPM helps quantify both the internal and external characteristics of a property through regression techniques; however, this model has limitations, such as the inclusion of many explanatory variables, which leads to the aforementioned multicollinearity problems, as well as the inability to find non-linear relationships between different variables [37].

However, these limitations can be corrected through the use of machine learning algorithms, which are more flexible, do not have specific requirements regarding data distribution, and can find non-linear relationships. For this reason, many authors use such methods to resolve the previously mentioned issues of HPM [38]. Thus, it is concluded that machine learning methods are a good complement to HPM, and by combining both techniques, the results can achieve a higher degree of fit. In fact, in [12], HPMs are seen as a solid model for highlighting the characteristics that most influence housing prices, and these authors aim to address the non-linear relationships in hedonic prices by performing a robustness test with a machine learning model. Similarly, [39] compare HPMs with spatial analysis, finding that the latter model is a good complement to hedonic prices, as both models provide results with a higher level of significance than each model individually.

In summary, in these works, the hedonic methodology is fundamental for understanding the variation in housing prices based on the bundle of characteristics it possesses. If the characteristics of a property are known and its prices can be estimated, HPM proves to be very useful and provides an opportunity to estimate the value [40].

Regarding models based on tree algorithms, such as decision trees, random forest, and gradient-boosted trees, these are the models most frequently selected as the most significant in the articles of this review. These models offer advantages in terms of performance with large datasets [41]. In his article, [42] confirms that the decision tree model is the one that best fits housing price prediction, both in terms of achieving the highest R^2 and the lowest error level. This study suggests that, most of the time, more complex models are given more consideration, but simpler models should also be observed more closely, as they can yield very similar results.

The random forest model is the one that best fits price prediction in most articles that use it, compared to other models. In studies like those of [31,43,44], the random forest model showed very significant results for housing price prediction and performed better than the other models, both in training and test data.

However, in the studies of [32,35,36], when comparing random forest with HPM and finding the former with better results, the authors confirm that random forest adequately reflects non-linearity and the real complexity of real estate markets and that, therefore, the random forest model is a good complement to HPM. Similarly, studies like [45] justify that the best solution is an ensemble classifier model, i.e., the combination of decision tree with random forest and gradient-boosted tree, as it shows excellent results.

Other studies confirm the good performance of tree-based models, such as those used to predict the housing price index in the US, where the application of a random forest model achieved an error margin of only 5% [46]. Or studies showing the effectiveness of random forest compared to multiple regression models in the housing market in Australia [47].

Regarding neural networks, these stand out in studies comparing them with HPM, as evidenced in the works of [2,27,30,34]. As mentioned earlier, neural networks provide more precise results than HPM in predicting housing prices. Moreover, in studies predicting housing prices in large cities, neural networks have been proven effective for this task, such as in the case of Taranto, Italy [48]. However, there have also been studies where neural networks were not effective, such as in the prediction of housing prices in Iran, where the fuzzy regression model proved more effective than neural networks [49].

The other models chosen by authors as the best models in their studies, such as XGBoost or LightGBM, were selected for providing better results in parameters and being more adaptable to the model.

Similarly, within the housing market and urban planning, artificial neural network models, random forests, and support vector machines are frequently used in land use classification aspects and are even employed to simulate land use planning processes [50].

It is important to note that many authors choose the models to develop based on the type of variables they want to include in the model, considering them significant for housing price prediction. Most of them use variables related to the external characteristics of the property, especially the location, such as neighborhood characteristics, environmental context, or the perception of the area and neighborhood held by potential buyers.

6. Conclusions

Access to housing is closely linked to its affordability, which is determined by the characteristics and environment of its location. The determination and understanding of the prices surrounding a property are fundamental aspects when it comes to accessing housing. This is why this systematic review was conducted, with the goal of gathering information on the set of internal and external characteristics that contribute to determining the price of a property.

After conducting the systematic review to understand the extent of studies focused on price prediction in the real estate sector, it has been found that advanced and machine learning models are frequently used in this sector. The importance of hedonic regressions in property valuation is emphasized, as this type of model contributes significantly to understanding how internal and external variables of a property affect the overall price and how it varies depending on each of these factors. It is also concluded that combining hedonic price models (HPM) with machine learning models results in more meaningful outcomes for property price prediction. Therefore, it is advisable to merge this advanced method with a machine learning model.

Machine learning-based prediction models, inherent to artificial intelligence (AI), provide an objective viewpoint and can offer strong results in predicting real estate prices. The use of these models, particularly considering external variables of properties (such as proximity to points of interest, neighborhood, environmental context, etc.), can deliver results that closely reflect real market prices with minimal margin of error. As such, these techniques can be a useful complement for real estate professionals, as well as for governments or public institutions in developing economic or housing policies.

The use of predictive algorithms with macro data can analyze variables that were previously not measured, providing clarity when establishing public policies, although they must be used with great caution. By combining new technologies with human capital, research in the real estate market field would be more fluid and yield better results in large cities. In this way, all market participants could understand the value of each property attribute and estimate the price fluctuations of each attribute separately, as well as the overall value of the property. This would help avoid the implementation of prices that diverge from the market reality, which are set using traditional valuation and prediction methods, and thus, have a high degree of subjectivity.

By analyzing the various studies on this topic, it has been found that much more research has been conducted on these prediction models in the real estate sector during the second decade of the 21st century. This leads to the conclusion that in the coming years, a wide variety of studies will be carried out with different samples, time periods, and advanced prediction models in this field. These studies will introduce new ideas and add data to help make this market more transparent, balanced, and accessible to everyone.

Author Contributions: Conceptualization, I.M.-F.; M.-T.S.-M. and M.P.-E.; methodology, I.M.-F.; M.-T.S.-M. and M.P.-E.; investigation, I.M.-F.; M.-T.S.-M. and M.P.-E.; writing—original draft preparation, I.M.-F.; M.-T.S.-M. and M.P.-E.; writing—review and editing, I.M.-F.; M.-T.S.-M. and M.P.-E.;

visualization, I.M.-F.; supervision, I.M.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Description of the Selected Studies

Table A1. Comparative studies of advanced price prediction models.

Authors	Compared Prediction Models	Summary/Conclusions
Hoxha, 2024 [42] Analysis period: 2019–2023 Sample size: 1512	Linear Regression, Decision Tree, K-Nearest Neighbors, and Support Vector Regression.	The prediction of housing prices in the city of Prishtina is analyzed through a dataset that includes housing prices and their characteristics, by comparing four machine learning models to extend the resulting knowledge to other real estate markets. The model that provides the best ability to predict prices is the Decision Tree model, as it achieves the lowest MSE and the highest R^2.
Soltani et al. 2022 [47] Analysis period: 1984–2016 Sample size: 428,000	Linear Regression, Decision Tree, Random Forest, and Gradient-Boosted Tree.	They analyze how housing prices in the metropolitan area of Adelaide vary based on the impact of certain characteristics, in this case, adding a spatial variable. It is demonstrated that non-linear models perform better than linear ones, and that applying a spatio-temporal variable is effective in machine learning models. The model with the best performance is the Gradient Boosted Tree, achieving the highest R^2 and the lowest RMSE and MAE.
Ligus and Peternek, 2017 [39] Analysis period: 2012–2014 Sample size: 6318	HPM and spatial analysis.	Understanding the preferences of homebuyers based on the environmental characteristics of the properties. By applying two econometric models, it is found that geographic characteristics are significant and cause variation in prices. The union of both models together presents more significant results than the HPM alone.
Chen et al. 2023 [51] Analysis period: 2014–2017 Sample size: 25,135	Gradient-Boosted Tree, Random Forest, Elastic Net, Lasso Regression, and Ridge Regression.	It aims to address the need to include uncommon variables in housing price variation, such as the effects and characteristics of neighborhoods, as well as the economic and ethnic conditions within them. As a result, transportation and the economic and sociodemographic characteristics of the neighborhood are significant, but the economic ones stand out as the most significant.

Table A1. *Cont.*

Authors	Compared Prediction Models	Summary/Conclusions
Lahmiri et al. 2023 [52] Analysis period: 2012–2013 Sample size: 414	Ensemble Regression Trees, Support Vector Regression, and Gaussian Regression.	By applying these three models, it aims to address the problems of housing price prediction. Ensemble regression trees achieve the best results for prediction as they exhibit the smallest error rate and low error variability, indicated by the range of the distribution. This makes them strong candidates for future predictions in Taiwan.
Wang, 2023 [2] Analysis period: 2015–2020 Sample size: 34,447	HPM, Neural Network, Lasso Regression, Ridge Regression, Regression Trees, Random Forest, and Gradient Boosting Machine.	The Google Street View tool is used to analyze the environmental location factors in housing prices through a neural network model. As a result, this model improves accuracy and provides more significant results than the others, this is because it recognizes edges and more complex shapes in its layers and highlights important features.
Chen et al. 2022 [31] Analysis period: 2013–2019 Sample size: 137,132	HPM, Random Forest, and Gradient Boosting Machine.	Using images of points of interest and neighborhood characteristics, this study examines how housing prices vary based on the image of the property compared to its intrinsic features. In general, the R^2 improves across all models when images are introduced, but the random forest model exhibits the smallest error rate, making it the one that delivers the most signif-icant results.
Hong et al. 2020 [32] Analysis period: 2006–2017 Sample size: 16,601	HPM and random forest.	It investigates the Random Forest model as a predictor of housing prices and compares it with HPM to assess the results. Between the two models, Random Forest proves to be more adaptable to the reality of the real estate market, as it achieves a higher R^2 and a lower MAPE, random forest is proposed as an excellent complement to the HPM.
Chen et al. 2017 [33] Analysis period: 2007–2010 Sample size: 3.991	HPM and Support Vector Machine.	It uses the Support Vector Machine method to predict housing prices, incorporating housing variables in the hedonic model to verify price variations. The support vector machine model demonstrates strong predictive power in forecasting housing prices with high accuracy.

Table A1. *Cont.*

Authors	Compared Prediction Models	Summary/Conclusions
Selim, 2009 [34] Analysis period: 2004 Sample size: 5.741	HPM and Artificial Neural Networks.	It analyzes the determinants of housing prices, including location. It compares both methods and finds that neural networks are the better alternative for price prediction as they are more accurate, although the HPM heteroscedasticity correction results in most variables being highly significant.
Ho et al. 2020 [43] Analysis period: 1996–2004 Sample size: 39,554	Random Forest, Support Vector Machine, and Gradient Boosting Machine.	It uses these three models for housing price evaluation. After applying the models, random forest and gradient boosting machine performed better than support vector machine (higher R^2 and lower error rate). However, the latter is still considered a very useful algorithm, providing accurate predictions within a strict time constraint.
Begum et al. 2022 [44] Analysis period: 2015–2019 Sample size: 506	Linear regression, decision tree and random forest.	It analyzes housing price prediction using three models, both advanced and machine learning-based. After analyzing the three models, random forest is the method that yields the best results, as it has the lowest error rate and performed well on both the training and test data.
Yoo et al. 2012 [35] Analysis period: 2000 Sample size: 4469	HPM, Linear Regression, Random Forest and Cubist.	Its goal is to use machine learning models for selecting hedonic variables and for housing sale price models. Several models were applied, and random forest resulted in the best accuracy in terms of modeling, with the potential to be useful for selecting important variables for the hedonic price equation.
Ceh et al. 2018 [36] Analysis period: 2008–2013 Sample size: 7407	HPM and random forest.	A comparison of both models, including the structural characteristics of the property, envi-ronmental information, and neighborhood characteristics, was conducted. The best predic-tions were obtained with the random forest method, which achieved a higher R^2 and a significantly lower error rate.
Tochaiwat and Pultawee, 2024 [45] Analysis period: 2011/2014/2017/2021 Sample size: 59	Decision Tree, Random Forest and Gradient-Boosted Tree.	After using multiple machine learning techniques to analyze urban development projects, it is found that combining models yields better results providing more accurate and realistic statistics compared to the individual models on their own.

Table A1. *Cont.*

Authors	Compared Prediction Models	Summary/Conclusions
Paik et al. 2023 [53] Analysis period: 2021–2022 Sample size: 58,342	Linear Regression, Decision Tree, Random Forest, LightGBM, Lasso Regression, Ridge Regression, Elastic Net, and XGBoost.	It aims to analyze the impact of metro stations and social capital on housing prices. To accomplish this, it compares eight machine learning methods to provide more information for determining housing prices. The LightGBM model has the smallest relative error between the actual and predicted values, and also performed better in terms of absolute error compared to the other models, making it the most suitable model for this study.
Rui and Liu, 2019 [54] Analysis period: 2010–2017 Sample size: 664	Neural Network (Short-Term Memory), GA (Genetic Algorithm), and Support Vector Regression.	To understand housing price fluctuations, a model composed of two methodologies is proposed as an experiment to test its effectiveness. The result shows that combining the neural network with genetic algorithms successfully predicts housing prices with a better feature selection process. However, a limitation is that if the dataset is small, the model weakens.
Kou et al. 2021 [55] Analysis period: 2018 Sample size: 158,888	Neural Networks, XGBoost, and Support Vector Machine.	The housing price valuation study is expanded by including regional clusters such as proxim-ity to workplaces or shopping centers. XGBoost presents the highest R^2, and furthermore, when economic clusters are added, they make the fit more accurate than all the traditional features.
Taecharungroj, 2021 [56] Analysis period: 2021 Sample size: 152,512	Random forest and XGBoost.	It aims to analyze the amenities of neighborhoods and condominium prices in the city. Both the popularity and availability of services drive condominium prices in non-linear ways. The XGBoost model shows a higher level of fit (R^2), making it perform better than random forest.
Chou et al. 2022 [27] Analysis period: 2013–2017 Sample size: 209,402	Artificial Neural Networks, Linear Regression, Regression Tree, Support Vector Machine, and Hybrid Model.	Four machine learning models are developed to predict housing prices. Additionally, a hybrid model is created for the same purpose. Neural networks achieved the best perfor-mance in R^2, RMSE, MAE, and MAPE. However, the hybrid model, by combining three models, demonstrates greater precision than each individual model, as it leverages the advantages of each.

Table A1. *Cont.*

Authors	Compared Prediction Models	Summary/Conclusions
Iban, 2022 [57] Analysis period: 2021 Sample size: 1002	Random forest, XGBoost, LightGBM and Gradient Boosting.	It aims to investigate the determinants considered in the models when valuing housing prices through the application of four machine learning methods. In this study, the Gradient Boosting model achieved the highest R^2 score. However, the XGBoost model presented the lowest MAPE and RMSE values, indicating better performance.
Núñez-Tabales et al. 2012 [30] Analysis period: 2006 Sample size: 2888	HPM and Artificial Neural Networks.	The goal is to obtain the implicit prices of housing characteristics by comparing the two models being analyzed. Neural networks demonstrate greater predictive power and more satisfactory results due to a higher degree of fit (R^2) and lower RMSE, MAE, and residual standard deviation rates.
Rico-Juan and Taltavull, 2021 [12] Analysis period: 2004–2012 Sample size: 392,412	K-Nearest Neighbors, Decision Tree, Random Forest, AdaBoost, XGBoost, CatBoost, Artificial Neural Networks (Multilayer), Linear Regression, Ridge Regression, Lasso Regression, and HPM.	By using HPM and several machine learning models, the aim is to determine which of them provides the most significant results in predicting housing prices. After analyzing all the models, the random forest model proves to be the most suitable for the task, showing a higher degree of fit and the lowest error rates.

Source: own elaboration.

References

1. Mark, J.; Kim, K. Segmentation of the housing market and its determinants: Seoul and its neighbouring new towns in Korea. *Aust. Geogr.* **2007**, *36*, 221–232.
2. Wang, G. The effect of environment on housing prices: Evidence from the Google Street View. *J. Forecast.* **2023**, *42*, 288–311. [CrossRef]
3. Bartholomew, K.; Edwing, R. Hedonic price effects of pedestrian- and transit-oriented development. *J. Plan. Lit.* **2011**, *26*, 18–34. [CrossRef]
4. Howard, G.; Liebersohn, J. Regional divergence and house prices. *Rev. Econ. Dyn.* **2023**, *49*, 312–350. [CrossRef]
5. Jayantha, W.M.; Oladinrin, O.T. Bibliometric analysis of hedonic price model using CiteSpace. *Int. J. Hous. Mark. Anal.* **2019**, *13*, 357–371. [CrossRef]
6. Xiao, Y. *Urban Morphology and Housing Market*; Springer: Berlin/Heidelberg, Germany, 2017.
7. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing, Long Beach, CA, USA, 4–9 December 2017; pp. 4766–4777.
8. Fan, C.; Cui, Z.; Zhong, X. House prices prediction with machine learning algorithms. In Proceedings of the ICMLC'18: 2018 10th International Conference on Machine Learning and Computing, Macau, China, 26–28 February 2018; pp. 6–10.
9. Rouhiaine, L. *Inteligencia Artificial: 101 Cosas Que Debes Saber Hoy Sobre Nuestro Future*; Editorial Planeta S.A.: Barcelona, Spain, 2018.
10. Barzegar, R.; Adamowski, J.; Moghaddam, A.A. Application of wavelet-artificial intelligence hybrid models for water quality prediction: A case study in Aji-Chai River, Iran. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 1797–1819. [CrossRef]
11. Glaeser, E.L.; Kominers, S.D.; Luca, M.; Naik, N. Big data and big cities: The promises and limitations of improved measures of urban life. *Econ. Inq.* **2016**, *56*, 114–137. [CrossRef]
12. Rico-Juan, J.R.; de la Paz, P.T. Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain. *Expert Syst. Appl.* **2021**, *171*, 114590. [CrossRef]
13. Goodman, A.C. Andrew Court and the invention of hedonic price analysis. *J. Urban Econ.* **1998**, *44*, 291–298. [CrossRef]

14. Rosen, S. Hedonic prices and implicit markets: Product differentiation in pure competition. *J. Political Econ.* **1974**, *82*, 34–55. [CrossRef]
15. Diewert, W.E.; de Haan, J.; Hendricks, R. Hedonic regressions and the decomposition of a house price index into land and structure components. *Econom. Rev.* **2011**, *34*, 106–126. [CrossRef]
16. Hwang, M.; Quigley, J. Economic fundamentals in local housing markets: Evidence from U.S. metropolitan regions. *J. Reg. Sci.* **2006**, *46*, 425–453. [CrossRef]
17. Mendoza, N.J. La Red de Hopfield. *Rev. Inf. Tecnol. Y Soc.* **2009**, *35*.
18. Shinde, P.P.; Shah, S. A review of machine learning and deep learning applications. In Proceedings of the International Fourth Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018; pp. 1–6.
19. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
20. Angra, S.; Ahuja, S. Machine learning and its application: A review. In Proceedings of the International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, India, 23–25 March 2017; pp. 57–60.
21. Song, Y.; Ma, X. Exploration of intelligent housing price forecasting based on the anchoring effect. *Neural Comput. Appl.* **2024**, *36*, 2201–2214. [CrossRef]
22. Arias-Odón, F. Investigación documental, investigación bibliométrica y revisiones sistemáticas. *Rev. Electrónica De Humanidades Educ. Y Comun. Soc.* **2023**, *31*, 9–28.
23. Manterola, C.; Astudillo, P.; Arias, E.; Claros, N.; Mincir, G. Revisiones sistemáticas de la literatura. Qué se debe saber acerca de ellas. *Cirugía Española* **2013**, *91*, 149–155. [CrossRef]
24. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. Declaración PRISMA 2020: Una guía actualizada para la publicación de revisiones sistemáticas. *Rev. Española De Cardiol.* **2021**, *74*, 790–799. [CrossRef]
25. Beltrán, Ó.A. Revisiones sistemáticas de la literatura. *Rev. Colomb. De Gastroenterol.* **2005**, *20*, 60–69.
26. Sarkis-Onofre, R.; Catalá-López, F.; Aromataris, E.; Lockwood, C. How to properly use the PRISMA statement. *Syst. Rev.* **2021**, *10*, 117. [CrossRef]
27. Chou, J.S.; Fleshman, D.B.; Truong, D.N. Comparison of machine learning models to provide preliminary forecasts of real estate prices. *J. Hous. Built Environ.* **2022**, *37*, 2079–2114. [CrossRef]
28. Rigatti, S.J. Random forest. *J. Insur. Med.* **2017**, *47*, 31–39. [CrossRef] [PubMed]
29. Shmueli, G. To explain or to predict? *Stat. Sci.* **2010**, *25*, 289–310. [CrossRef]
30. Tabales, J.M.; y Ocerin, J.M.; Villamandos, N.C.; Carmona, F.J. Obtención de precios implícitos para atributos determinantes en la valoración de una vivienda. *Rev. Int. De Adm. Y Finanz.* **2012**, *5*, 41–54.
31. Chen, M.; Liu, Y.; Arribas-Bel, D.; Singleton, A. Assessing the value of user-generated images of urban surroundings for house price estimation. *Landsc. Urban Plan.* **2022**, *226*, 104486. [CrossRef]
32. Hong, J.; Choi, H.; Kim, W.S. A house price valuation based on the random forest approach: The mass appraisal of residential property in South Korea. *Int. J. Strateg. Prop. Manag.* **2020**, *24*, 1–13. [CrossRef]
33. Chen, J.-H.; Ong, C.F.; Zheng, L.; Hsu, S.-C. Forecasting spatial dynamics of the housing market using Support Vector Machine. *Int. J. Strateg. Prop. Manag.* **2017**, *21*, 273–283. [CrossRef]
34. Selim, H. Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Syst. Appl.* **2009**, *36*, 2843–2852. [CrossRef]
35. Yoo, S.; Im, J.; Wagner, J.E. Variable selection for hedonic model using machine learning approaches: A case study in Onondaga County, NY. *Landsc. Urban Plan.* **2012**, *107*, 293–306. [CrossRef]
36. Čeh, M.; Kilibarda, M.; Lisec, A.; Bajat, B. Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *Int. J. Geo-Inf.* **2018**, *7*, 168. [CrossRef]
37. Schläpfer, F.; Waltert, F.; Segura, L.; Kienast, F. Valuation of landscape amenities: A hedonic pricing analysis of housing rents in urban, suburban, and periurban Switzerland. *Landsc. Urban Plan.* **2015**, *141*, 24–40. [CrossRef]
38. Wang, D.; Li, V.J. Mass appraisal models of real estate in the 21st century: A systematic literature review. *Sustainability* **2019**, *11*, 7006. [CrossRef]
39. Ligus, M.; Peternek, P. *Impacts of Urban Environmental Attributes on Residential Housing Price in Warsaw (Poland): Spatial Hedonic Analysis of City Districts*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 155–164.
40. Crespo, R.; Grêt-Regamey, A. Local hedonic house-price modelling for urban planners: Advantages of using local regression techniques. *Environ. Plan. B Plan. Des.* **2013**, *40*, 664–682. [CrossRef]
41. Afonso, B.; Melo, L.; Oliveira, W.; Sousa, S.; Berton, L. Housing prices prediction with a deep learning and random forest ensemble. In Proceedings of the Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), Salvador, Brasil, 15–18 October 2019; pp. 389–400.
42. Hoxha, V. Comparative analysis of machine learning models in predicting housing prices: A case study of Prishtina's real estate market. *Int. J. Hous. Mark. Anal.* **2024**. [CrossRef]

43. Ho, W.K.O.; Tang, B.-S.; Wong, S.W. Predicting property prices with machine learning algorithms. *J. Prop. Res.* **2020**, *38*, 48–70. [CrossRef]
44. Begum, A.; Kheya, N.J.; Rahman, M.Z. Housing price prediction with machine learning. *Int. J. Innov. Technol. Explor. Eng.* **2022**, *11*, 42–46. [CrossRef]
45. Tochaiwat, K.; Pultawee, P. House type specification for housing development project using machine learning techniques: A study from Bangkok Metropolitan Region, Thailand. *J. Environ. Plan.* **2024**, *23*, 403. [CrossRef]
46. Adetunji, A.B.; Akande, O.N.; Ajala, F.A.; Oyewo, O.; Akande, Y.F.; Oluwadara, G. House price prediction using random forest machine learning technique. *Procedia Comput. Sci.* **2022**, *199*, 806–813. [CrossRef]
47. Soltani, A.; Heydari, M.; Aghaei, F.; Pettit, C.J. Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. *Cities* **2022**, *131*, 103941. [CrossRef]
48. Chiarazzo, V.; Caggiani, L.; Marinelli, M.; Ottomanelli, M. A neural network-based model for real estate price estimation considering environmental quality of property location. *Transp. Res. Procedia* **2014**, *3*, 810–817. [CrossRef]
49. Azadeh, A.; Sheikhalishahi, M.; Boostani, A. A flexible neuro-fuzzy approach for improvement of seasonal housing price estimation in uncertain and non-linear environments. *South Afr. J. Econ.* **2014**, *82*, 567–582. [CrossRef]
50. Chaturvedi, V.; de Vries, W.T. Machine learning algorithms for urban land use planning: A review. *Urban Sci.* **2021**, *5*, 68. [CrossRef]
51. Chen, Y.; Jiao, J.; Farahi, A. Disparities in affecting factors of housing price: A machine learning approach to the effects of housing status, public transit and density factors on single-family housing price. *Cities* **2023**, *140*, 104432. [CrossRef]
52. Lahmiri, S.; Bekiros, S.; Avdoulas, C. A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization. *Decis. Anal. J.* **2023**, *6*, 100166. [CrossRef]
53. Paik, J.; Baek, S.-J.; Kim, J.-W.; Ko, K. Influence of social overhead capital facilities on housing prices using machine learning. *Appl. Sci.* **2023**, *13*, 10732. [CrossRef]
54. Rui, L.; Liu, L. Predicting housing price in China based on long short-term memory incorporating modified genetic algorithm. *Soft Comput.* **2019**, *23*, 11829–11838.
55. Kou, J.; Du, J.; Fu, X.; Zhang, G.Z.; Wang, H.; Zhang, Y. *The Effect of Regional Economic Clusters on Housing Price*; Lecture Notes in Artificial Intelligence; Springer: Berlin/Heidelberg, Germany, 2021.
56. Taecharungroj, V. Google Maps amenities and condominium prices: Investigating the effects and relationships using machine learning. *Habitat Int.* **2021**, *118*, 102409. [CrossRef]
57. Iban, M.C. An explainable model for the mass appraisal of residences: The application of tree-based machine learning algorithms and interpretation of value determinants. *Habitat Int.* **2022**, *128*, 102409. [CrossRef]