

Article

Machine Learning Insights: Exploring Key Factors Influencing Sale-to-List Ratio—Insights from SVM Classification and Recursive Feature Selection in the US Real Estate Market

Janusz Sobieraj ¹  and Dominik Metelski ^{2,*} 

¹ Department of Building Engineering, Warsaw University of Technology, 00-637 Warsaw, Poland; janusz.sobieraj@pw.edu.pl

² Department of International and Spanish Economics, University of Granada, 18071 Granada, Spain

* Correspondence: dominik@correo.ugr.es

Abstract: The US real estate market is a complex ecosystem influenced by multiple factors, making it critical for stakeholders to understand its dynamics. This study uses Zillow Econ (monthly) data from January 2018 to October 2023 across 100 major regions gathered through Metropolitan Statistical Area (MSA) and advanced machine learning techniques, including radial kernel Support Vector Machines (SVMs), used to predict the sale-to-list ratio, a key metric that indicates the market health and competitiveness of the US real estate. Recursive Feature Elimination (RFE) is used to identify influential variables that provide insight into market dynamics. Results show that SVM achieves approximately 85% accuracy, with temporal indicators such as Days to Pending and Days to Close, pricing dynamics such as Listing Price Cut and Share of Listings with Price Cut, and rental market conditions captured by the Zillow Observed Rent Index (ZORI) emerging as critical factors influencing the sale-to-list ratio. The comparison between SVM alphas and RFE highlights the importance of time, price, and rental market indicators in understanding market trends. This study underscores the interplay between these variables and provides actionable insights for stakeholders. By contextualizing the findings within the existing literature, this study emphasizes the importance of considering multiple factors in housing market analysis. Recommendations include using pricing dynamics and rental market conditions to inform pricing strategies and negotiation tactics. This study adds to the body of knowledge in real estate research and provides a foundation for informed decision-making in the ever-evolving real estate landscape.

Keywords: Support Vector Machines (SVM); Recursive Feature Elimination (RFE); real estate forecasting; housing market dynamics; US real estate; predictive modeling; sale-to-list ratio; panel data analysis; Zillow econ



Citation: Sobieraj, J.; Metelski, D. Machine Learning Insights: Exploring Key Factors Influencing Sale-to-List Ratio—Insights from SVM Classification and Recursive Feature Selection in the US Real Estate Market. *Buildings* **2024**, *14*, 1471. <https://doi.org/10.3390/buildings14051471>

Academic Editors: Yaning Qiao, Hongping Yuan, Zeyu Wang and Xiaozhi Ma

Received: 5 March 2024

Revised: 16 April 2024

Accepted: 16 May 2024

Published: 18 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The real estate market is a complex and multifaceted domain, influenced by a myriad of factors that shape its evolution over time [1–3]. Understanding the dynamics of this market is of paramount importance, not only for academic research, but also for professionals, investors, and policymakers who seek to navigate this landscape effectively [4]. One of the key concepts that is critical to understanding the dynamics of the housing market is the concept of market tightness [5–7]. This reflects the balance between supply and demand dynamics within a given real estate market, capturing the ratio of buyers to sellers. For example, a study by Anenberg and Ringo [8] addresses the role of market tightness, defined as the ratio of buyers to sellers, in driving short-term dynamics in housing sales and prices. A tighter market, characterized by a higher ratio of buyers to sellers, indicates greater competition among buyers and more leverage for sellers, leading to higher prices and faster sales. Conversely, a slacker market with a lower ratio of buyers to sellers indicates less demand relative to supply, potentially resulting in slower sales and more price flexibility.

The sale-to-list ratio serves as a valuable proxy for measuring market tightness, reflecting the balance of power between buyers and sellers [9]. By analyzing the trends and drivers of this specific metric, researchers and policymakers can gain valuable insights into the underlying dynamics of the housing market, such as the relative influence of supply and demand factors. Furthermore, understanding market tightness through the lens of the sale-to-list ratio is critical to developing a comprehensive understanding of housing market conditions and informing effective policy decisions.

More specifically, the sale-to-list ratio holds significance for several reasons [9–13]. Firstly, it serves as a pivotal gauge of housing market health, reflecting the percentage of listing prices realized in sales. Secondly, it aids both buyers and sellers in price determination and gauging market competitiveness. Thirdly, it facilitates the forecasting of future market trends, encompassing changes in home values and inventory levels. Finally, it complements other market indicators, such as for-sale inventory and (new construction) sales count, offering a comprehensive market overview [9]. It is important to note that the sale-to-list ratio has garnered attention for its utility in assessing market health and trends, guiding real estate decisions for investors, professionals, and policymakers. Rising ratios may signal investment opportunities, while declining ratios may warrant cautious strategies [7]. Previous studies have utilized this metric to evaluate perceived investment risks and bargaining power dynamics, offering insights into market competitiveness and short-term price forecasting [5–7,14]. However, exploring the key factors influencing the sale-to-list ratio still remains a significant challenge amidst the intricate interplay of market dynamics [9–13]. Therefore, this study aims to address this knowledge gap by pursuing a dual objective: to provide a nuanced exploration of the influence of a variety of different predictor variables on the sale-to-list ratio, and to elucidate the underlying mechanisms driving the predictive power of the identified variables through advanced machine learning techniques such as Support Vector Machines (SVMs) and Recursive Feature Elimination (RFE). More specifically, the study addresses several specific research questions related to the topic of the study and the results of the hybrid approach, i.e., they relate to the confusion matrix accuracy scores, the alpha values, and the Recursive Feature Elimination (RFE). For example, we want to know what the predictive accuracy of the SVM model is in classifying sale-to-list ratios, and how does this accuracy affect decision-making processes and practical implications for real estate stakeholders such as buyers, sellers, and investors in terms of pricing strategies, negotiation tactics, and overall market insights? How do the alpha values derived from the SVM model in the study contribute to an understanding of the relationships between the predictor variables and the sale-to-list ratio in the U.S. real estate market, and what actionable insights do they provide for stakeholders? Specifically, how can this understanding inform strategic decision-making processes for real estate professionals, investors, and policymakers? Finally, what is the quantitative relationship between the variables identified by Recursive Feature Elimination (RFE) as the most influential factors in predicting the sale-to-list ratio, and how do these variables collectively contribute to understanding market dynamics in the U.S. real estate market?

There is another feature of the real estate market that needs to be mentioned here. Because the real estate landscape is a dynamic and complex system influenced by a variety of factors, including economic fluctuations [15], demographic shifts [16], and regional disparities [17], this dynamic environment poses significant challenges to researchers and policymakers seeking to understand and address the complexities of housing markets. To address these complex, nonlinear problems, which often involve large data sets, there is a growing need to apply sophisticated analytical techniques, such as machine learning and hybrid scientific methods. Such advanced approaches have the potential to provide deeper insights and more accurate predictions than traditional econometric models, which can struggle to capture the nuanced relationships and patterns inherent in housing market data [18]. By leveraging the capabilities of machine learning algorithms and hybrid approaches, researchers can gain a more complete understanding of the drivers and dynamics that shape housing markets [19–26]. For example, the study by Michele et al. [24] uses

machine learning techniques to address the problem of duplicate ads in online housing listings, which can skew the analysis of housing supply and demand. The authors show that by correcting for this bias, online listing data can become a powerful tool for the real-time analysis of housing markets. The study by Ho et al. [25] explores the use of various machine learning algorithms, such as gradient boosting, random forest, and Support Vector Machines, to predict house prices. The authors show that these advanced techniques can achieve very accurate predictions of house prices, outperforming traditional econometric models. The study by Alzain et al. [26] applies machine learning algorithms, including a stochastic gradient descent-based support vector random forest and gradient boosting machine, to forecast real estate prices in Saudi Arabia. The study shows that these AI-based methods can achieve highly accurate results. By the same token, this study contributes to this discourse by using advanced machine learning techniques, specifically SVM classification and RFE, to identify influential variables that affect the sale-to-list ratio. The use of SVM classification provides a nuanced understanding of housing market dynamics, while RFE streamlines model complexity, improving interpretability and predictive accuracy. Through meticulous data pre-processing and the use of comprehensive data sets spanning 70 months across major metropolitan areas in the United States, this study attempts to explore the intricate dynamics of U.S. real estate, with a particular focus on better explaining and predicting the sale-to-list ratio. In particular, recent trends, such as US homes selling below asking prices in December 2022, juxtaposed with instances of sale-to-list ratios exceeding 100 in the second quarter of 2022, underscore the volatility of the market [10–13].

Overall, the study seeks to examine the predictive power of several real estate metrics, including temporal indicators such as days pending and days to close, pricing dynamics such as listing price cuts and share of listings with price cuts, and the rental market conditions as captured by the Zillow Observed Rent Index (ZORI). Using advanced machine learning techniques such as SVMs and RFE, the study aims to identify which of these variables is most influential in predicting the sale-to-list ratio. In addition, the study quantifies the impact of each variable on the accuracy of the predictive model, providing insight into the relative importance of different market indicators. Furthermore, the research seeks to explore how these insights could inform strategic decision-making for various stakeholders in the U.S. real estate market, including sellers, buyers, investors, and policymakers. Finally, the study aims to contribute to a deeper understanding of the complex dynamics of the U.S. real estate market and to provide actionable insights for informed decision-making in this sector.

The paper is organized as follows. Section 2 addresses the existing literature on the topic, providing a thorough review of the relevant theoretical and empirical studies. Section 3 outlines the materials and methods used in the study. This includes a detailed description of the SVM model and classifier, as well as the Recursive Feature Elimination (RFE) method used for feature selection, and a description of the variables and data employed in the study. Section 4 presents the results of the study, including the confusion matrix, accuracy scores, RFE results, performance over subset size, and alpha values (reflecting the predictor variables' importance) for the SVM model with the radial kernel function. In Section 5, the study provides a comprehensive discussion of the findings, unraveling the nuanced relationships among the identified variables and their broader implications for real estate investment. Finally, Section 6 presents the study's conclusions and offers recommendations for future research, emphasizing the ongoing need for dynamic models that can adapt to the evolving landscape of the real estate market.

2. Literature Review

The study of the sale-to-list ratio, which reflects the relationship between the final sale price and the original list price, is an important but understudied aspect of the real estate market. While there is a scarcity of literature directly examining the sale-to-list ratio, several studies have explored related concepts such as housing market tightness [8,27–31],

the balance of power between buyers and sellers [8,28,31,32], and the dynamics of supply and demand [33]. One of the few studies that directly examines the sale-to-list ratio is the work by Zhang et al. [9] which establishes a robust theoretical framework and provides empirical evidence linking the sale-to-list ratio to housing values and prices. The authors use lagged Pearson correlation tests, Granger causality tests, and cointegration tests to demonstrate the causal relationships between the sale-to-list ratio, for-sale inventory, sale count nowcast, and both nominal and real housing values.

Given the limited literature on the sale-to-list ratio, researchers often turn to proxies such as housing market tightness, the ratio of buyers to sellers, and housing market liquidity or even housing market “hotness” [5,31] in different contexts. For example, Anenberg and Ringo [8] discuss how market tightness, defined as the ratio of buyers to sellers, plays a crucial role in explaining short-run housing dynamics. They find that a tight market, as evidenced by the rapid sale of recent listings, informs sellers that they can raise prices, while slower sales indicate the need for price reductions. The study by Ngai and Tenreiro [31] further highlights the importance of market tightness, noting that a “hot” market is characterized by high prices, more buyers and sellers, and a larger number of transactions. The authors argue that a model accounting for both market tightness and thickness (the number of buyers and sellers) better explains the positive correlation between prices and transaction volume. Carrillo [5] also emphasizes the role of listing price strategies in influencing buyer behavior and transaction outcomes, thus impacting housing market “hotness” (which also can be viewed as a proxy for the sale-to-list ratio). The study constructs a “housing market hotness” index based on list prices, sale prices, and time on the market, capturing the degree of competition and liquidity in the housing market. Similarly, Anenberg and Ringo [28] stress the importance of measures of market tightness for understanding the underlying dynamics of the housing market. They note that these measures can serve as useful proxies for the balance between housing supply and demand, which is crucial for policymakers and industry stakeholders.

To better understand the factors influencing the sale-to-list ratio, this study aims to identify and substantiate their relationship with the variable under study. These factors have been analyzed in the context of market tightness, and some may serve as early indicators to inform decision-making. A complete list of the identified indicators and the justification for their use in explaining and understanding the sale-to-list ratio is presented in Section 3.3—Variables and data. In general, the drivers of the sale-to-list ratio can be categorized into (1) price-related indicators such as list price (LP), listing price cut (LPC), new construction sale price (NCSP), sale price (SP), Zillow Home Value Index (ZHVI); (2) inventory-related indicators such as for-sale inventory (FSI), and new construction sales count (NCSC); (3) time-related indicators such as days to pending (DTP), and days to close (DTC); and (4) miscellaneous indicators such as share of listings with a price cut (SLPC), MSA size rank (SR), and Zillow Observed Rent Index (ZORI). Each of these variables can be related in some way to housing market tightness, which supports their association with the sales-to-list ratio. For example, the study by Yilmaz et al. [27] found that higher rental market liquidity, a measure closely related to the Zillow Observed Rent Index (ZORI), is associated with lower housing market tightness, suggesting a buyer’s market (a proxy for the sale-to-list ratio). The study by Bich et al. [34] suggests that list price strategies are likely to play a crucial role in explaining market tightness, i.e., they may influence the sale-to-list ratio (the study does not explicitly address the sale-to-list ratio). The study by Leamer [35] shows that new home sales volumes exhibit a clear cyclical pattern, with significant declines during recessions, while house prices do not exhibit the same pronounced cyclicity. This suggests that the tightness of the housing market, as reflected in sales volumes, is a more important driver of new home prices than other factors.

A key set of variables are the price-related indicators. The initial list price (LP) directly influences price dynamics and the potential bargaining range, as shown by the work of Bich et al. [34] and the theoretical model developed by Ngai and Sheedy [29]. Listing price cuts, captured by the Listing Price Cut (LPC) variable, reflect market responsiveness,

pricing accuracy, and seller motivation, as shown by Gabrovski and Ortego-Martí [30], and are associated with housing market tightness. The pricing of new construction, as measured by the new construction sales price (NCSP), may also affect overall market pricing, with Leamer's [35] research suggesting that housing market tightness is a more important driver of new construction prices than other factors. The final sales price (SP) reflects market valuation and the accuracy of initial pricing, with the work of Gilbukh and Goldsmith-Pinkham [36] demonstrating how an agent's level of experience can serve as a proxy for the overall tightness or looseness of the housing market.

Time-related indicators, such as days to pending (DTP) and days to close (DTC), provide insights into market demand and the efficiency of the sales process. Gabrovski and Ortego-Martí's [30] sophisticated search-and-matching framework directly relates the time on the market variable to housing market tightness, while Anenberg and Ringo's [8] and Sklarz's [37] research suggests that slower sales, indicating a less tight market, would inform sellers that they may need to lower prices to achieve a timely sale.

Inventory-related variables, including for-sale inventory (FSI) and new construction sales count (NCSC), reflect the balance between supply and demand, which is a key driver of market tightness. Ngai and Sheedy's [29] theoretical framework linking housing market dynamics to the concept of market tightness provides an indirect rationale for the link between market tightness and for-sale inventory, while Anenberg and Ringo's [8] work provides a framework for understanding how housing supply can affect overall housing market tightness. By analyzing this comprehensive set of variables from the Zillow Econ Database, researchers can gain a deeper understanding of the factors that shape the sale-to-list ratio and, consequently, overall housing market tightness.

Finally, the choice of the SVMs as the primary method in our study, to relate these variables to the sale-to-list ratio, stems from its robustness in handling high-dimensional data and its ability to effectively classify complex datasets. SVMs are well suited for our research objective of predicting the sale-to-list ratio in the U.S. real estate market because of their flexibility in handling nonlinear relationships and their ability to identify intricate patterns within the data. In addition, SVMs have the advantage of maximizing the margin between different classes, which improves the generalization performance of the model. Furthermore, SVMs have been widely used in real estate research due to their ability to handle large data sets and provide accurate predictions. Given the complexity of real estate market dynamics and the need for accurate predictions, SVMs prove to be a suitable choice for our study. Moreover, the comparison between SVM alphas and Recursive Feature Elimination (RFE) underscores the effectiveness of SVMs in capturing the underlying patterns and influential variables in the housing market, further justifying our methodological approach. In addition, the study uses the hybrid SVM-RFE method simply because it can better handle variables that can produce some sort of statistical problems, such as multicollinearity or heterogeneity (for example), and since variables such as list price, sale price, and Zillow Home Value Index (ZHVI) are all dollar denominated and likely to be highly correlated, conventional regression analysis is not an option here. The use of Support Vector Machine Recursive Feature Elimination (SVM-RFE) can be beneficial in situations with variables that exhibit statistical problems such as multicollinearity or heterogeneity. And while the specific variables mentioned do not necessarily require the use of SVM-RFE, it is an effective feature selection method that can handle such problems better than traditional regression analysis. While traditional regression analysis can still be a viable option, provided that multicollinearity is properly addressed through techniques such as ridge regression, lasso regression, or principal component regression, the key advantage of SVM-RFE is its ability to handle nonlinear relationships and complex feature interactions that may not be easily captured by standard regression methods. Thus, while SVM-RFE can be a useful tool, it is not necessarily required in all cases with correlated variables, and conventional regression analysis remains a valid option if multicollinearity is properly accounted for.

In summary, while the existing literature directly examining the sale-to-list ratio is limited, researchers have explored related concepts that provide valuable insights into this important but understudied aspect of the real estate market. These studies have investigated housing market tightness, the balance of power between buyers and sellers, and the dynamics of supply and demand. They highlight the complex interplay of various factors, such as market conditions, buyer and seller behavior, and the dynamics of supply and demand. By synthesizing insights from these theoretical perspectives, researchers can develop a more comprehensive understanding of the factors that influence the sale-to-list ratio.

3. Materials and Methods

3.1. Support Vector Machines

Support Vector Machines (SVMs) is a nonparametric statistical learning method that has recently been used to solve various scientific problems and has found many applications in many fields. To understand what SVM is, one must first be familiarized with the kernel function which can be expressed as: $k(x, x') = \langle \Phi(x), \Phi(x') \rangle$ [38]. It is a type of implicit mapping Φ from input data to multi-dimensional feature spaces. SVMs require user-defined parameters, and each parameter has a different effect on kernels, so the accuracy of SVM classification depends on the choice of parameters and kernels. Simply put, a kernel function is a function that projects the images of two data points x, x' into a feature space (such projection can be characterized as $Q : X \rightarrow H$) and then returns an inner product between the images of these two points. The learning process is then performed in the feature space, while the data points appear only in the context of a dot product $\langle \Phi(x), \Phi(x') \rangle$ with other points, which is often referred to as the “kernel trick” [38]. It turns out that from the perspective of memory constraints, using kernel functions is a computationally much more efficient solution than projecting x and x' onto the feature space H . For kernel-based systems, it is only important to choose a suitable kernel function, which then allows the user to deal with a multidimensional space and to use quadratic programming to solve complex problems. Employing SVM enables the selection of an appropriate kernel function for high-dimensional feature spaces, thereby resolving complex research issues without incurring excessive computational expenses. Notably, this approach avoids directly transforming input data into multidimensional feature spaces through mapping, which typically does not involve extracting new features or altering the original data structure. Consequently, maintaining the integrity of the initial dataset becomes crucial during tasks like classification, regression, and outlier detection, for which the SVM method is used. According to Shin et al. [39], SVM has an advantage over other similar pattern recognition methods such as Backpropagation Neural Networks (BNNs) in that it is able to extract the correct solution with a small training set, which is undoubtedly a major advantage of this method. This is because SVMs are able to capture certain geometric features of the feature space without extracting the network weights from the training data. To train SVMs, quadratic optimization is used. However, the problem is that even with a small dataset, training an SVM based on solving quadratic problems faces a quadratic programming constraint (QP) [40], which can be quite a problem. For m training points, this would mean that the corresponding calculation would have to be performed for an $m \times m$ matrix. Due to the limited ability to deal with the size of the problem, this would limit the application potential of the SVM method [41]. Fortunately, there are iterative methods to solve this problem, including chunking, Sequential Minimal Optimization (SMO) and simple SVM algorithms [40,42–44]. These methods rely on appropriate scaling and allow for relatively simple computations [41].

In the realm of scientific inquiry, SVMs represent a vector-based learning approach rooted in statistical learning theory [45]. Originating in the late 1960s and evolving over many decades [45–47], SVMs have matured into a sophisticated scientific method. They serve as a type of learning algorithm employed for the estimation of multidimensional functions [45]. Initially, SVMs were primarily a theoretical tool for analyzing function esti-

mation problems, tailored to specific datasets. Vapnik's [45] seminal work delves into both the theoretical underpinnings and algorithmic intricacies of SVMs, outlining generalization conditions and algorithmic strategies for function estimation challenges. Distinguished by its broad applicability compared to traditional statistical frameworks, SVMs leverage a straightforward linear method within a high-dimensional feature space, introducing non-linear relationships with the input space. Notably, SVM's elegance lies in its avoidance of explicit calculations within this multi-dimensional feature space. This approach empowers the resolution of diverse scientific and practical problems like classification, regression, novelty detection, and feature reduction. Figure 1 illustrates how SVMs facilitate hypothesis testing concerning the target space assumption, approximation error, and generalization error related to modeling inaccuracies. Specifically, the SVM hypothesis quantifies the distance between data points and decision boundaries or lines, offering a robust framework for scientific analysis and inference.

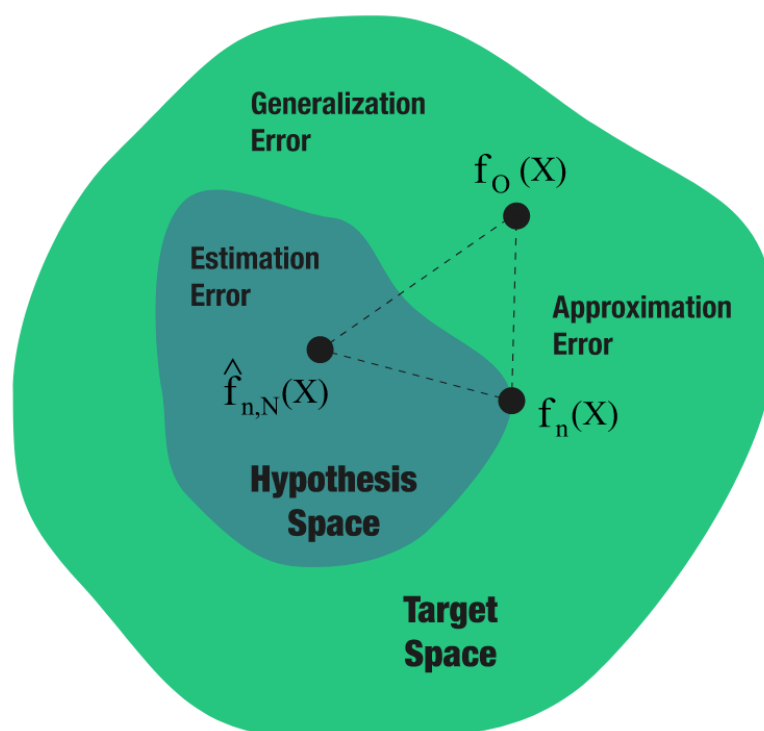


Figure 1. Modeling Errors. Source: own elaboration based on [48].

By selecting appropriate kernel functions and underlying algorithms, the structural adaptability of SVMs can be tailored to the specific tasks to which the method is applied [38]. Unlike some alternative approaches, SVMs operate without incorporating posterior probabilities, instead relying on the robust theoretical framework of statistical learning theory. This foundational approach provides a rigorous yet practical basis for using SVMs to address complex and challenging engineering problems. Many of the fundamentals of SVM methods are explained in the study by Salcedo-Sanz et al. [49], which discusses kernel theory, SVM fundamentals, support vector regression (SVR), SVM applications in signal processing, and the fusion of metaheuristic techniques with SVMs. The versatility of SVMs allows to address many real-world problems in various engineering domains, resulting in numerous successes. In addition, Salcedo-Sanz et al. [49] explain how SVMs excel at handling multidimensional, heterogeneous, and disorganized data sets, extracting valuable insights to inform targeted solutions and applications.

3.2. Recursive Feature Elimination (RFE) and the Hybrid SVM-RFE Approach

Recursive Feature Elimination (RFE) is an algorithm utilized for feature selection in machine learning. Initially, RFE considers all features within the training dataset and employs a machine learning algorithm to rank their importance. Subsequently, it iteratively eliminates the least important features until a specified number of features remain selected [50]. In the realm of SVM classification, RFE proves valuable for pinpointing the most pertinent features for the model. The mathematical representation of RFE in SVM classification involves: (1) calculating the weight of each feature through the SVM algorithm and arranging them based on importance [51]; (2) selecting and removing the feature with the highest weight iteratively until the desired number of features is reached [51]; (3) training the SVM model on the chosen features [51]. Research indicates that employing RFE in SVM classification enhances the classification accuracy compared to using SVM alone [52]. This improvement stems from RFE's ability to reduce dataset dimensionality, aiding the SVM algorithm in recognizing underlying patterns more effectively and enhancing its performance. In conclusion, RFE stands out as a potent tool for feature selection in SVM classification tasks, facilitating the identification of critical features that can significantly boost classification performance.

The hybrid approach used in this study can be presented as follows. We assume that the SVM classifier is formulated as the following optimization problem:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (1)$$

subject to:

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, n \quad (2)$$

where \mathbf{w} is the weight vector, b is the bias term, $\xi = \{\xi_1, \xi_2, \dots, \xi_n\}$ are the slack variables, and C is the regularization parameter.

The RFE algorithm for SVM classification is then described as follows: (1) the SVM model is first trained using the entire feature set and the weight vector \mathbf{w} is obtained; (2) the ranking score for each feature j as $|w_j|$, where w_j is the j -th element of \mathbf{w} is computed; (3) the features with the smallest ranking score (i.e., the least important features) are then removed, and finally; (4) steps 1–3 are repeated until the desired number of features is reached.

3.3. Variables and Data

The study utilizes data sourced from the Zillow Econ database, covering a period of 70 months from January 2018 to October 2023. The dataset comprises 100 major US regions selected based on size rank by Metropolitan Statistical Area (MSA) and contains 13 variables, including the dependent variable (sale-to-list ratio), providing a comprehensive overview of the US real estate landscape. The data was collected from [<https://www.zillow.com/research/data>, accessed on 5 February 2024] and transformed into a panel data frame. The selection of variables is a meticulous process guided by domain knowledge and a comprehensive understanding of real estate dynamics. Each variable is selected for its logical association with the sales-to-list ratio, providing unique insights into the forces driving the housing market conditions. For example: (1) "size rank" reflects the relative size of the MSA region, influencing supply and demand dynamics; (2) "Sale Price" indicates the average sale price, a critical factor influencing the sale-to-list ratio; (3) "Days to Pending" addresses the time it takes for a property to go from listing to pending, a measure of market activity; (4) "Days to Close" represents the average time from listing to closing, contributing to forecast accuracy; (5) Zillow Rental Index illuminates rental market conditions, influencing housing demand. Moreover, the selected variables capture various aspects of the housing market, such as temporal indicators like "Days to Pending" and "Days to Close", and price dynamic indicators like "Listing Price Cut" and "Share of Listings with Price Cut". Economic metrics such as the Zillow Home Value Index (ZHVI) and Zillow Observed Rent Index (ZORI) are also included. These variables were chosen

based on their potential utility and relevance in predicting the sale-to-list ratio, the primary focus of this study. The complete list of variables is detailed in Table 1.

Table 1. Description of the variables used in the study.

Variable	Units	Description	Substantiation
Sale-to-List ratio (STL)	Percentage (%)	The sale-to-list ratio is a real estate metric that measures the final sale price of a home compared to its original listing price, expressed as a percentage. It can be viewed as a proxy for housing market tightness. Specifically, the sale-to-list ratio is calculated by dividing the final sale price by the initial asking price and expressing it as a percentage.	Anenberg and Ringo [8] discuss how market tightness, or the ratio of buyers to sellers, plays an important role in explaining short-run housing dynamics. Ngai and Tenreiro [31] highlight the importance of market tightness, noting that “a hot market is one with high prices, more buyers and sellers, and an unambiguously larger number of transactions”.
Days to Pending (DTP)	Days	The number of days between listing a property and accepting an offer. This metric is used to understand the speed of the housing market and the time it takes for homes to go under contract [53,54].	Useful in predicting the sale-to-list ratio as it reflects the time it takes for a property to attract a buyer, which can indicate market demand and pricing accuracy. Gabrovski and Ortego-Martí's [30] paper provides a sophisticated search-and-matching framework that directly relates the time on the market variable to housing market tightness and other key market conditions. This offers an alternative perspective to the more granular Zillow metrics (i.e., days to pending), while still capturing the overall dynamics of the housing sales process.
Days to Close (DTC)	Days	The number of days between accepting an offer and closing the sale [53].	Useful in predicting the sale-to-list ratio as it reflects the efficiency of the sales process and the accuracy of initial pricing. Anenberg and Ringo [8] argue that slower sales, indicating that the market is not so tight, would inform them that they may need to lower prices to make a timely sale. This suggests that days on the market (or time to sell) is closely tied to the tightness of the housing market. Sklarz [37] directly discusses the relationship between days on market and housing market tightness.
For-Sale Inventory (FSI)	Number of homes	The number of properties available for sale [9].	Ngai and Sheedy [29] appear to have established a theoretical framework linking housing market dynamics to the concept of market tightness. The listing rate nt is measured as the ratio of new listings Nt to the stock of owner-occupied houses not already for sale, that is, $nt = Nt / (Kt - Ut)$ [where nt : The listing rate, which is the ratio of new listings (Nt) to the stock of owner-occupied houses not already for sale ($Kt - Ut$); Nt : The number of new listings; Kt : The total stock of owner-occupied houses; Ut : The number of houses already for sale]. This provides an indirect rationale for the association of the market tightness and for-sale inventory (FSI).

Table 1. Cont.

Variable	Units	Description	Substantiation
List Price (LP)	US Dollars (\$)	The initial price at which a property is listed for sale [55].	Useful in predicting the sale-to-list ratio as it directly influences the pricing dynamics and potential negotiation range. Bich et al.'s [34] study suggests that list price strategies likely play a crucial role in explaining market tightness, meaning that it potentially affects sale-to-list ratio (the study does not address the sale-to-list ratio explicitly). Ngai and Sheedy's [29] study developed a theoretical model that examines how the listing rate, which is related to the ratio of buyers to sellers (i.e., market tightness), impacts the housing search and matching process.
Listing Price Cut (LPC)	US Dollars (\$)	The reduction in the listing price [55].	Useful in predicting the sale-to-list ratio as it reflects market responsiveness and pricing accuracy, and can indicate seller motivation and market conditions. Gabrovski and Ortego-Marti [30] provide evidence linking housing market tightness to housing price cuts and reductions. Gabrovski and Ortego-Marti [30] argue that "higher house prices drive entry of investors and lowers market tightness—or alternatively, they post more vacancies for any given number of buyers". This indicates that as housing market tightness decreases, with more vacancies relative to buyers, sellers are more likely to offer price cuts or reductions to attract buyers.
New Constr. Sale Price (NCSP)	US Dollars (\$)	The price at which a newly constructed property is sold.	Useful in predicting the sale-to-list ratio as it reflects the pricing dynamics and demand for new construction, which can impact overall market pricing. The study by Leamer [35] examines the relationship between housing starts, sales volumes, and prices. It shows that the sales volumes of new homes exhibit a clear cyclical pattern, with substantial dips during recessions, while home prices do not exhibit the same pronounced cyclicity. This suggests that housing market tightness, as reflected in sales volumes, is a more important driver of new construction prices than other factors.
New Constr. Sales Count (NCSC)	Number of homes	The number of newly constructed properties sold.	Useful in predicting the sale-to-list ratio as it indicates market activity and demand for new construction, which can impact overall market dynamics and pricing. The study by Anenberg and Ringo [8] provides a framework for understanding how the supply of homes (which NCSC would influence) can affect the overall tightness of the housing market and, in turn, housing prices and sales. The study indicates that NCSC, as a component of overall housing supply, would be expected to influence market tightness and, consequently, housing market outcomes.

Table 1. Cont.

Variable	Units	Description	Substantiation
Sale Price (SP)	US Dollars (\$)	The price at which a property is sold.	Useful in predicting the sale-to-list ratio as it reflects the final market valuation and the accuracy of initial pricing and negotiation dynamics. The study by Gilbukh and Goldsmith-Pinkham [36] provides evidence linking housing sale prices to market tightness. It basically shows that an agent's work experience is a key determinant of their ability to successfully sell homes, with more experienced agents having a significant advantage in sale probability compared to less experienced agents. Importantly, this experience gap varies significantly over the housing cycle—it is wider during market downturns and narrower during booms. This indicates that the distribution of agent experience in the market is a reflection of broader housing market conditions. When there are more experienced agents, it signals a tighter market where homes sell more quickly. Conversely, a predominance of inexperienced agents points to a looser market with slower home sales. In this way, an agent's experience level serves as a proxy for the overall tightness or looseness of the housing market, providing insight into the balance of supply and demand.
Share of Listings With a Price Cut (SLPC)	Percentage (%)	The proportion of listings that have undergone a price cut [56].	Useful in predicting the sale-to-list ratio as it reflects market conditions, seller flexibility, and pricing accuracy, and can indicate buyer negotiation power. The study by Gabrovski and Ortego-Marti [30] provides relevant evidence linking housing market tightness to the share of listings with price cuts (SLPC). The key points from that study are that higher house prices drive the increased entry of investors, which lowers market tightness by leading sellers to post more vacancies relative to the number of buyers. When market tightness decreases, with more vacancies compared to buyers, sellers become more likely to offer price cuts or reductions in order to attract buyers. This suggests that as housing market tightness decreases, with an oversupply of listings relative to buyer demand, sellers will be more inclined to cut listing prices in an effort to generate sales. The share of listings with price cuts can therefore be seen as an indicator of decreasing market tightness. By this logic, the share of listings with price cuts (SLPC) can be used as a proxy to infer changes in housing market tightness. When SLPC increases, it signals a loosening of the market and reduced tightness, as sellers compete for a limited pool of buyers by offering price reductions.

Table 1. Cont.

Variable	Units	Description	Substantiation
Size Rank (SR)	Ordinal ranking (1st, 2nd, 3rd, etc.)	The ranking of the MSA region compared to other MSA regions.	Useful in predicting the sale-to-list ratio as it reflects the size of a given MSA and its impact on pricing and market demand. Prior evidence has shown that there is a relationship between the sale-to-list ratio and MSA's rank in that in larger, more populated housing markets, there is often a greater imbalance between housing supply and demand, leading to more competition among buyers and higher sale-to-list ratios [57,58].
Zillow Home Value Index (ZHVI)	US Dollars (\$)	The Zillow Home Value Index, which measures the typical value of homes in a given area.	Useful in predicting the sale-to-list ratio as it provides a benchmark for property valuation and market trends. Kotova and Zhang [59] use the ZHVI as a measure of home prices in their analysis. The researchers state that the ZHVI is "the price line in the year plots" and that their results are similar if they use other home price indices like the CoreLogic price index. This indicates that the ZHVI is a reliable proxy for tracking overall home price trends and movements in the housing market. Home prices are a key indicator of housing market tightness—higher prices generally signal a tighter market with lower inventory and higher demand. By using the ZHVI as a measure of home prices, Kotova and Zhang's [59] study is indirectly linking it to housing market tightness. The ZHVI serves as a representation of the price dynamics in the housing market, which are closely tied to the balance between supply and demand, and overall market conditions.
Zillow Observed Rent Index (ZORI)	US Dollars (\$)	The Zillow Observed Rent Index, which measures the typical rent in a given area.	Useful in predicting the sale-to-list ratio as it reflects rental market dynamics and can indicate property investment potential and market demand. Kotova and Zhang's [59] paper also uses the Zillow Rent Index (ZRI) as a measure of rents, which are another important factor influencing housing market tightness. The relationship between home prices, rents, and other housing market variables examined in the paper further supports the idea that the ZHVI can be used as an indicator of market tightness. The article by Yilmaz et al. [27] found that higher rental market liquidity (an area closely related to ZORI) is associated with lower housing market tightness, suggesting a buyer's market (a proxy for sale-to-list ratio).

Source: own elaboration.

The sale-to-list ratio, a key metric studied in this research, compares the final sale price of a property to its initial asking price, expressed as a percentage [10–13]. This ratio provides valuable insights into negotiating trends, market competitiveness, and buyer–seller dynamics. For instance, a ratio above 100% indicates a potential seller's market, while a ratio below 100% suggests a potential buyer's market [60,61]. It is important to note that the average home in the US sold for several percent below its asking price in December

2022, as a result of the housing market slowing. Just a few months before that, in the second quarter of 2022, the sale-to-list price ratio went above 100 (as shown in Figures 2 and 3; note that the sale-to-list ratio can be expressed not only as a percentage, but also in decimal form). This reflected the high housing demand and the need of prospective home buyers to bid above the asking price. Housing demand—as measured in pending home sales—went up, as mortgage rates were historically low and plummeted once rates were increased.

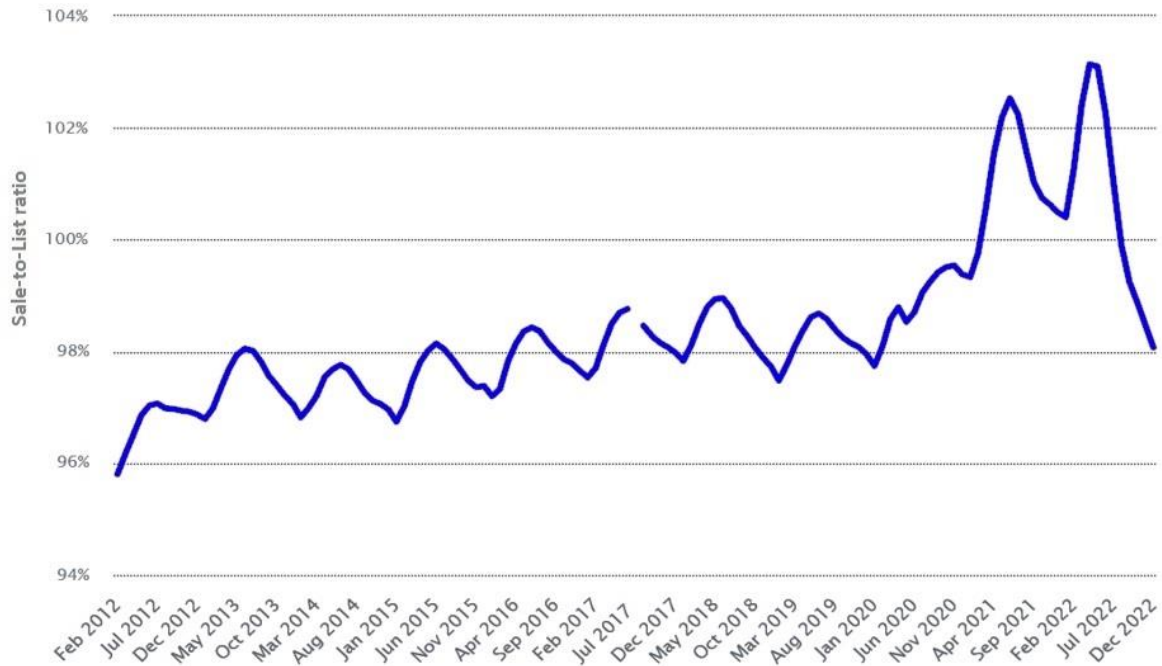


Figure 2. Sale-to-list price ratio of housing sales in the US 2012–2022; Source: own elaboration based on data from [statista.com](https://www.statista.com).

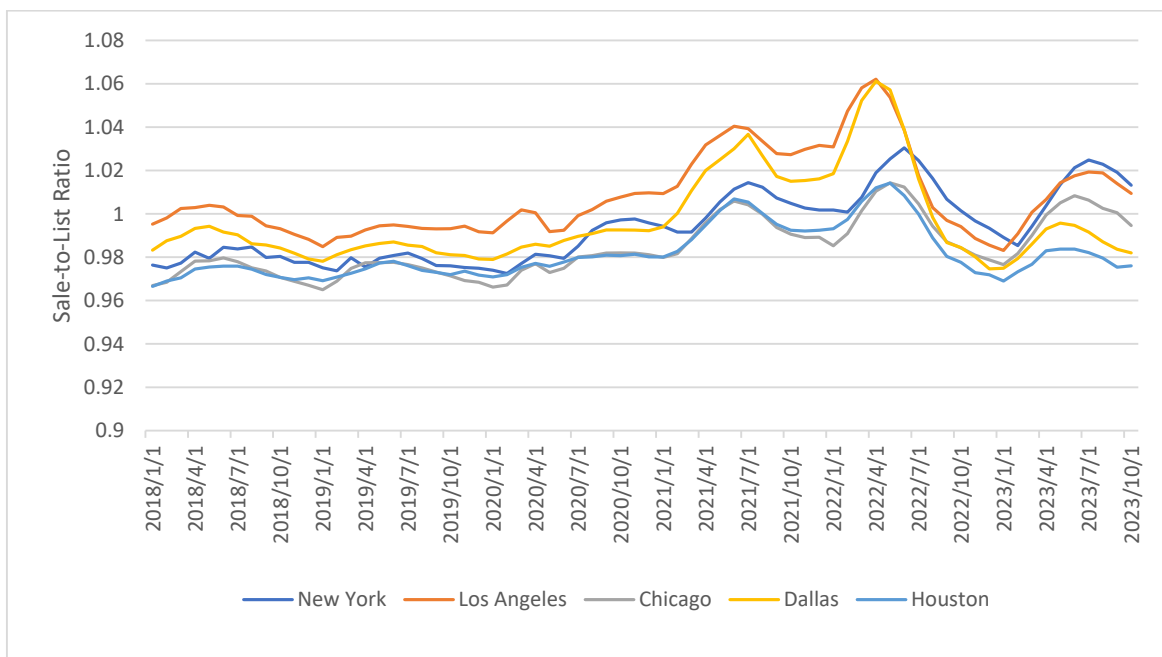


Figure 3. Sale-to-list price ratio across the five biggest 5 MSAs over the period 2018–2023; Source: own elaboration based on data from Zillow Econ Database. Note: the biggest MSA (with respect to its size rank) is New York, followed by Los Angeles, Chicago, Dallas, and Houston.

Table 2 provides a statistical summary of the data [i.e., variables used in the study including the studied variable sale-to-list ratio (STL)].

Table 2. Statistical summary of the data (variables) used in the study.

Vars	n	Mean	sd	Median	Trimmed	Mad	Min	Max	Range	Skew	Kurtosis	se
STL	6783	1.00	0.02	0.99	0.99	0.02	0.89	1.18	0.29	1.52	4.85	0.00
SR	7000	51.26	29.37406	51.5	51.225	37.8063	1	102	101	−0.00328	−1.19273	0.351087
SP	6961	306,462.2	175,654	255,000	275,301.2	100,075.5	95,000	1,550,000	1,455,000	2.688775	10.12511	2105.34
DTP	6853	39.6066	20.65162	36	37.43589	20.7564	8	154	146	1.052513	1.409263	0.249467
DTC	6773	36.20124	7.786966	35	35.48422	5.9304	9	72	63	1.059554	2.150149	0.094619
LP	6956	351,419.3	179,816.5	299,000	320,011.3	102,299.4	85,000	1,495,000	1,410,000	2.30421	7.069295	2156.005
LPC	6917	0.035423	0.008394	0.034245	0.03483	0.008081	0.016287	0.078596	0.062309	0.74521	0.746386	0.000101
SLPC	6964	0.184758	0.064211	0.179283	0.181576	0.063698	0.007576	0.488177	0.480601	0.544647	0.421007	0.000769
FSI	6952	7426.526	9963.77	4045.5	5257.462	3142.371	132	98,511	98,379	4.071141	22.45198	119.5003
ZORI	6990	1496.498	470.9466	1373.765	1428.191	373.0954	786.9547	3265.834	2478.879	1.355271	1.722398	5.632913
ZHVI	6858	315,443.9	182,042.5	262,792.3	283,041.6	104,144.7	103,085.5	1,521,706	1,418,621	2.527099	8.672603	2198.234
NCS	6636	267.4702	317.2863	160	203.7454	157.1556	5	2496	2491	2.775884	9.866544	3.894916
NCSP	5608	402.040	170,248.3	364,412.5	375,571.1	104,649.3	156,000	1,945,000	1,789,000	2.928937	13.14059	2273.416

Source: own elaboration; Note: STL—Sale-to-List Ratio, SR—Size Rank, SP—Sale Price, DTP—Days to Pending, DTC—Days to Close, LP—List Price, LPC—Listing Price Cut, SLPC—Share of Listing with a Price Cut, FSI—For-Sale Inventory, ZORI—Zillow Observed Rent Index, ZHVI—Zillow Home Value Index, NCSC—New Construction Sales Count, and NCSP—New Construction Sale Price.

The statistical summary of the data reveals insightful characteristics of the key variables utilized in the study. Firstly, the sale-to-list ratio (STL) indicates the degree of competitiveness in the housing market, with a mean value close to 1 across the 100 major regions studied. This suggests a balanced market between supply and demand. Additionally, the size rank (SR) highlights the relative scale of the regions, with a mean value indicating a moderate size distribution. The reason why the maximum value of SR is 102 and not 100 is that for two MSAs (with the size rank below 100) there was an excessive scarcity of datapoints (in that even an imputation made no sense in their case); therefore, we removed them and added two MSAs with the SR above 100, namely with the ranks 101 and 102. The variability in sale price (SP) and list price (LP) underscores the diverse pricing dynamics observed across different regions, with considerable variations in both mean and median values. Days to pending (DTP) and days to close (DTC) reflect the efficiency of property transactions, with mean values indicating a typical timeframe for real estate transactions to be completed. The frequency of listing price cuts (LPC) and the share of listings with a price cut (SLPC) provide insights into market competitiveness and pricing strategies, with varying degrees of variability observed. Moreover, the for-sale inventory (FSI) reflects the supply dynamics within each region, while the Zillow Observed Rent Index (ZORI) and Zillow Home Value Index (ZHVI) capture rental market conditions and home value trends, respectively. Finally, the variables new construction sales count (NCSC) and new construction sale price (NCSP) offer insights into new development activity and pricing trends within the real estate market. Overall, the statistical summary highlights the diverse and multifaceted nature of the variables under study, underscoring the complexity of the real estate landscape and the need for comprehensive analytical approaches to understand underlying market dynamics.

It is important to note that the sale-to-list ratio can provide valuable insights for both buyers and sellers, as well as real estate professionals. It can help identify negotiating trends, such as whether it is a buyer's market or a seller's market in a particular area. A relatively lower ratio compared to other neighborhoods or boroughs may indicate more negotiating power for buyers, while a higher ratio indicates greater competition among buyers and more leverage for sellers [9].

As for the data utilized in the SVM model and for RFE (the hybrid SVM-RFE approach), they underwent thorough cleaning and preprocessing, which involved addressing missing values and scaling. Prior to model development, a crucial step was taken to handle missing data, ensuring the completeness and reliability of the dataset. The missing values were imputed using the k-Nearest Neighbors (kNN) imputation method, implemented through the Visualization and Imputation of Missing Values (VIM) package in R. This technique

leverages the proximity of data points to estimate missing values, thereby enhancing the integrity of the dataset. Furthermore, the study made use of panel data, a longitudinal dataset that captures observations across multiple time periods and entities (in this case, MSAs). The employment of panel data offers several advantages in real estate research, including the ability to account for time dynamics and individual region trajectories, thereby contributing to the robustness of the results. Additionally, the data was scaled, a crucial step for enhancing model performance. The scaling of variables is essential for ensuring the effectiveness of the model, as demonstrated in Figure 4.

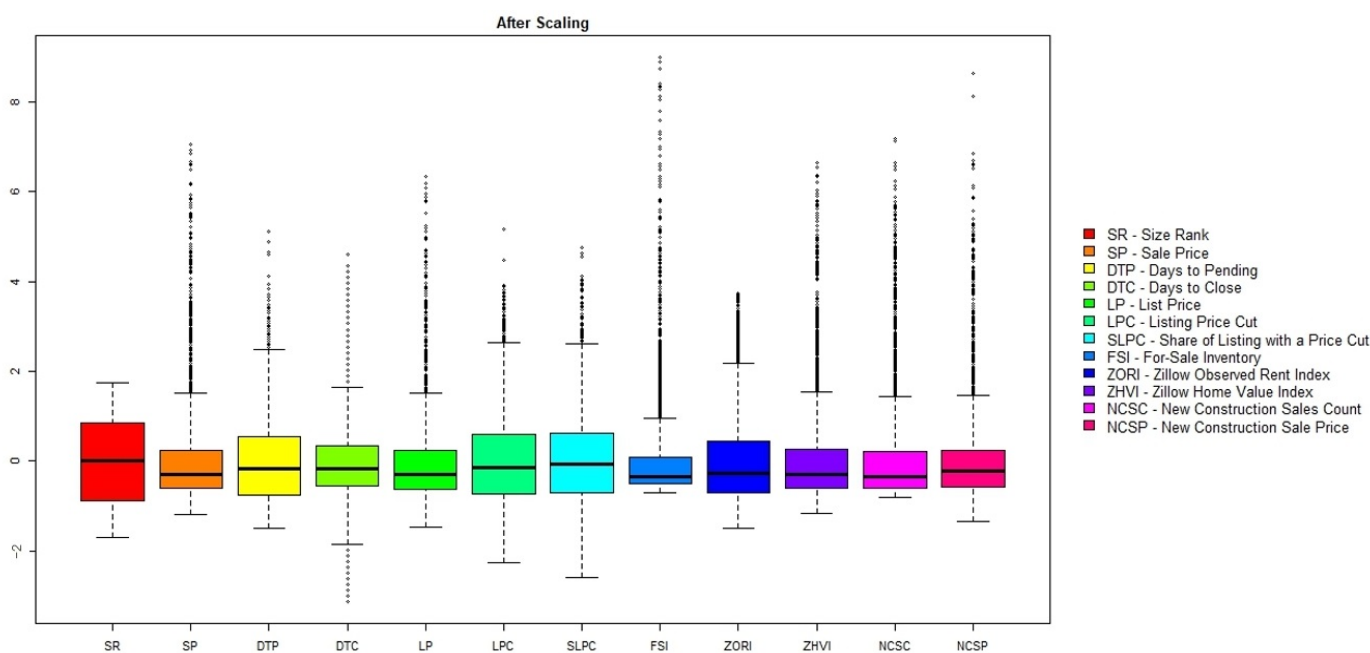


Figure 4. The data after scaling; source: own elaboration.

Variable scaling, a pivotal preprocessing step, exerts a notable influence on SVM model performance. By standardizing the range of values, scaling mitigates the dominance of individual variables in the modeling process, resulting in enhanced convergence speed and predictive accuracy. The subsequent improvement in model performance underscores the importance of meticulous preprocessing steps in real estate forecasting.

Overall, the methodology employed in this study combines SVM classification with Recursive Feature Elimination to forecast sale-to-list ratios. SVM, chosen for its ability to capture non-linear relationships, is complemented by RFE, which systematically identifies the most influential features. As mentioned earlier, variable selection is guided by domain knowledge and understanding of real estate dynamics, ensuring a robust analysis. SVM alphas provide additional insight into the relative importance of each variable, enhancing interpretability. More importantly, the panel data analysis accounts for time dynamics and regional variations, contributing to the reliability of results. In summary, the methodology integrates advanced analytical techniques with domain knowledge-driven variable selection to provide comprehensive insights into the US real estate market dynamics. Through meticulous data collection and analysis, this study aims to uncover new insights and contribute to the ongoing discourse in real estate research.

4. Results

The predictive model for the sale-to-list ratio is built with the use of the SVM classification algorithm. SVM, a supervised learning method, excels in classification tasks, especially in scenarios with complex decision boundaries and high-dimensional feature spaces [45]. Our implementation uses the 'e1071' package in R, which provides a flexible and efficient SVM framework. SVM works by identifying an optimal hyperplane that separates different

classes in the feature space. It maximizes the distance between classes, which improves the generalization ability of the model [41,48]. The choice of a radial kernel for the classification task (chosen based on a comparison between different kernel types, i.e., linear, polynomial, and radial) is chosen to give the best accuracy results, and tuning parameters such as cost are optimized through a cross-validated grid search, ensuring the robustness of the model. In turn, the RFE method served as a central component in distilling the extensive set of predictor variables. This feature selection technique systematically prunes less informative features, iteratively refining the model's predictive power [50–52]. The 'caret' package in R facilitated the seamless implementation of RFE, allowing us to identify the most influential variables. The RFE process involved systematically removing less relevant variables, evaluating the impact on model performance at each step. The variables that consistently contributed to improved accuracy were deemed critical for predicting sale-to-list ratio. This approach allowed us to pinpoint the most influential characteristics, providing a refined understanding of the factors that drive home values.

The results encompass an exploration of key metrics, an interpretation of confusion matrix results, an analysis of the RFE outcomes, an examination of SVM weights alphas, and an evaluation of scaling effects on model performance. They provide a comprehensive picture of the dynamics of the US real estate market and offer valuable insights for industry stakeholders.

Accuracy, a pivotal metric for assessing the SVM model's performance, stands at approximately 85% (as shown in Table 3), outperforming the 'No Information Rate' significantly with a p -value $<2.2 \times 10^{-16}$, indicating the model's effectiveness. The Kappa statistic of 0.774 suggests substantial agreement beyond chance.

Table 3. Accuracy scores.

Statistic	Value
Accuracy	0.8493
95% CI	(0.8295, 0.8676)
No Information Rate	0.4407
p -Value [Acc > NIR]	$<2.2 \times 10^{-16}$
Kappa	0.774
Mcnemar's Test p -Value	$<2.2 \times 10^{-16}$

Source: own elaboration.

The accuracy underscores the efficacy of the model in predicting the sale-to-list ratio. However, while accuracy is significant, a comprehensive evaluation of model performance necessitates complementing this metric with additional analyses. It is imperative to complement this score with additional metrics and analysis to ensure a thorough evaluation of the model performance.

The Confusion Matrix provides a granular view of the predictive accuracy of the SVM model (see Table 4 and Figure 5 below).

Table 4. Confusion Matrix.

Reference	Prediction	Frequency	Precision	Recall
Low	Low	443	0.948608	1
Low	Medium	24	0.961686	1
Low	High	0	0.663812	1
Medium	Low	24	0.948608	1
Medium	Medium	436	0.961686	1
Medium	High	157	0.663812	1
High	Low	0	0.948608	1
High	Medium	6	0.961686	1
High	High	310	0.663812	1

Source: own elaboration.

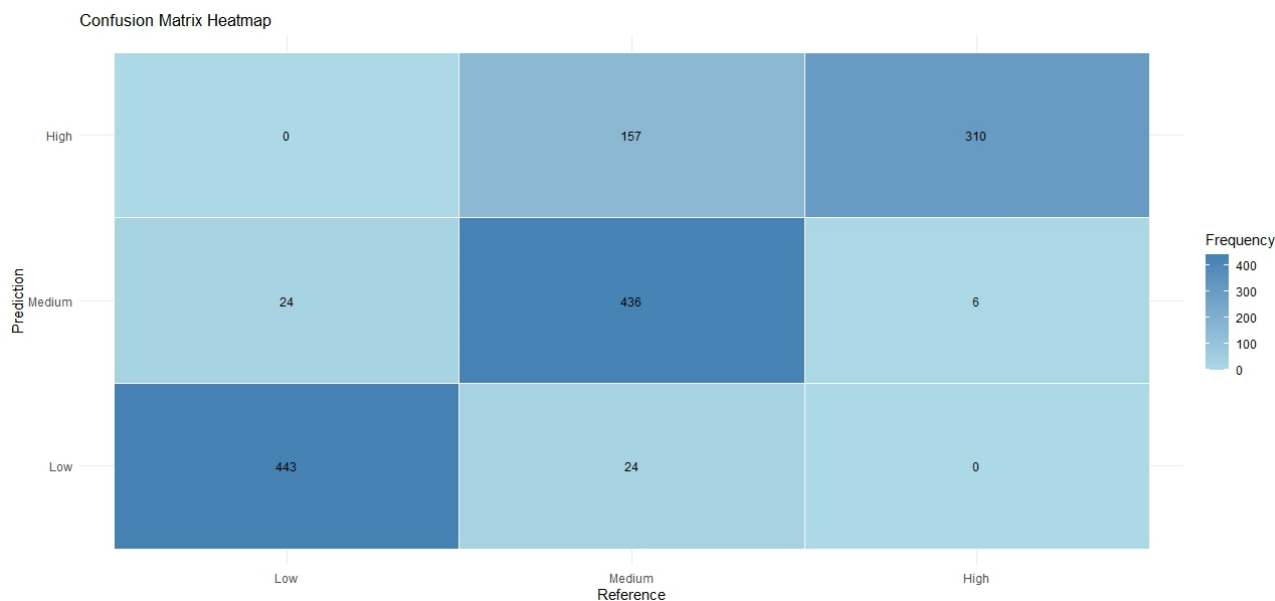


Figure 5. Confusion Matrix Heatmap.

Analyzing the confusion matrix reveals that the model performed well in predicting low and medium sale-to-list ratios, with high precision and recall values. However, for high ratios, while precision was lower, recall remained at 1, indicating that the model correctly identified all instances of high ratios. By delineating true positives, true negatives, false positives, and false negatives, the matrix enables an understanding of the model's strengths and weaknesses. Precision, recall, and F1 scores derived from the confusion matrix offer nuanced insights into the model's classification capabilities, ensuring a robust assessment of its predictive power.

It is important to note that the observed phenomenon of similar precision values for different combinations of reference and prediction categories in the confusion matrix may seem puzzling at first, but it can be explained by considering the distribution and characteristics of the data set under analysis. Precision as a metric evaluates the accuracy of a classifier in predicting the instances of a given class. In this context, the similar precision values across different reference and prediction combinations suggest that the classifier has consistent accuracy in correctly identifying the instances of the predicted class, relative to the instances it identifies overall. While the specific reasons for this observation may vary depending on the intricacies of the dataset and the underlying patterns captured by the classifier, it underscores the importance of interpreting evaluation metrics comprehensively and contextualizing them within the broader framework of the classification task at hand. Further analysis and exploration of additional evaluation metrics, such as recall, accuracy, and F1 score, provide a more holistic understanding of the classifier's performance and its implications for the given problem domain.

Examining Statistics by Class further highlights the model's performance across different classes (as shown in Table 5).

As shown in Table 4, sensitivity was notably high for low and high ratios, indicating the model's ability to detect these classes effectively. Specificity values were also strong across all classes, showing the model's capacity to correctly identify true negatives. Overall, the SVM model with RFE feature selection demonstrated robust predictive capabilities for sale-to-list ratios in real estate, particularly excelling in identifying low and medium ratios. The high accuracy, along with strong sensitivity and specificity values, underscores the model's reliability and potential practical application in real estate forecasting scenarios.

RFE outcomes yield critical insights into feature importance, with the stepwise elimination of variables shedding light on their impact on predictive accuracy. Notably, the top five features—Days to Pending, Listing Price Cut, Share of Listings with Price Cut, Days to

Close, and ZORI—emerge as the most influential factors in predicting the sale-to-list ratio (as shown in Table 6). This revelation underscores the significance of specific variables in shaping the dynamics of the US real estate market.

Table 5. Statistics by Class.

Class	Low	Medium	High
Sensitivity	0.9486	0.7066	0.981
Specificity	0.9743	0.9617	0.8552
Pos Pred Value	0.9486	0.9356	0.6638
Neg Pred Value	0.9743	0.8062	0.9936
Prevalence	0.3336	0.4407	0.2257
Detection Rate	0.3164	0.3114	0.2214
Detection Prevalence	0.3336	0.3329	0.3336
Balanced Accuracy	0.9614	0.8342	0.9181

Source: own elaboration.

Table 6. Recursive Feature Selection Results: Cross-Validated (10-fold) Performance Over Subset Size.

Variables	RMSE	Rsquared	MAE	RMSESD	RsquaredSD	MAESD	RFE
Days to Pending (DTP)	0.011863	0.535952	0.007721	0.001044	0.050047	0.000242	*
Listing Price Cut (LPC)	0.010833	0.613523	0.007126	0.000855	0.038162	0.000285	*
Share of Listings With a Price Cut (SLPC)	0.00956	0.69826	0.006382	0.000598	0.030582	0.000282	*
Days to Close (DTC)	0.008818	0.750748	0.005749	0.000592	0.024477	0.000204	*
Zillow Observed Rent Index (ZORI)	0.0082	0.792349	0.005257	0.000564	0.027883	0.000237	*
For-Sale Inventory (FSI)	0.007282	0.831117	0.004679	0.000427	0.02254	0.000216	
New Constr. Sales Count (NCSC)	0.006985	0.848399	0.004477	0.000446	0.017772	0.000179	
New Constr. Sale Price (NCSP)	0.006802	0.859017	0.004336	0.000439	0.014561	0.000136	
Sale Price (SP)	0.00654	0.866543	0.00416	0.000428	0.018364	0.000136	
Size Rank (SR)	0.006461	0.870519	0.004094	0.000454	0.015609	0.000164	
Zillow Home Value Index (ZHVI)	0.00637	0.875724	0.004027	0.000473	0.013964	0.000181	
List Price (LP)	0.006343	0.874604	0.004011	0.000457	0.014449	0.000168	

Source: own elaboration in R. Note: The table displays the performance metrics (RMSE, Rsquared, MAE) for different subset sizes during the recursive feature selection process. The top five variables selected, marked with an asterisk (*), are: Days to Pending, Listing Price Cut, Share of Listings with a Price Cut, Days to Close, and ZORI, emphasizing their importance in predicting the target variable.

Figure 6, illustrating RMSE, MAE, and Rsquared metrics, provides a comprehensive evaluation of the SVM model's performance. These metrics offer a nuanced assessment of predictive accuracy, capturing the model's ability to minimize errors and capture variability in the sale-to-list ratio. This holistic evaluation facilitates a balanced understanding of the model's strengths and areas for potential refinement.

The subset size measurement scale refers to the number of variables included in the model at each iteration of the RFE process. In the context of Figure 6, the subset size represents the incremental increase in the number of predictor variables considered in the SVM model (hence the X-axis is from 1 to 12). The RFE algorithm systematically evaluates subsets of variables, starting from a single variable and incrementally adding more variables until a predefined stopping criterion is met, such as reaching a maximum subset size or achieving optimal model performance. Each row in incremental value on the X-axis corresponds to a specific subset size, ranging from one variable to the total number of variables in the dataset (12). For example, the subset size equal to 1 represents the performance metrics obtained when considering only the single variable Days to Pending (DTC), while the last row represents the performance metrics obtained when considering all variables in the dataset. The subset size scale thus provides insight into how model performance varies with the increasing complexity of predictor variable combinations, allowing us to assess the impact of feature selection on predictive accuracy and model interpretability.

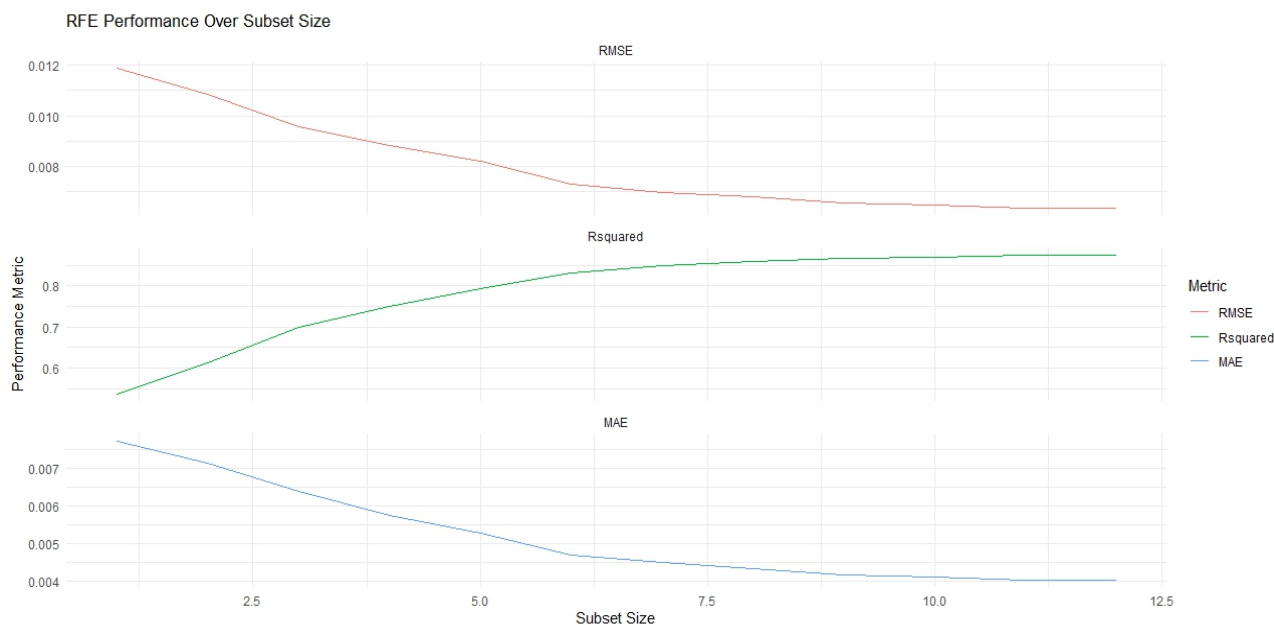


Figure 6. RFE Performance Over Subset Size; source: own elaboration in R.

The alphas, serving as SVM weights in the radial kernel, offer insights into the relative importance of each variable (as shown in Table 7 and Figure 7). A careful examination of these weights unveils the influence of variables in predicting the sale-to-list ratio, with variables possessing higher absolute alphas exerting a more significant impact on the model's decision-making process. This analysis provides a unique lens through which to understand the influence of variables on the target variable (i.e., the sale-to-list ratio). The results presented in Table 7 and Figure 7 provide insights into the relative importance of 12 key variables in predicting the sale-to-list ratio in the US real estate market using a radial kernel SVM model.

Table 7. Variable Importance—SVM (radial kernel).

SVM Model Feature	Alpha
Size Rank (SR)	−1.67285
Sale Price (SP)	0.558076
Days to Pending (DTP)	0.068493
Days to Close (DTC)	0.333333
List Price (LP)	0.592199
Listing Price Cut (LPC)	0.041028
Share of Listings With a Price Cut (SLPC)	−1.26482
For-Sale Inventory (FSI)	0.009794
Zillow Observed Rent Index (ZORI)	2.677687
Zillow Home Value Index (ZHVI)	3.039253
New Construction Sales Count (NCSC)	0.003348
New Construction Sale Price (NCSP)	0.795975

Source: own elaboration in R.

The alphas, serving as SVM weights, indicate the influence of each variable on the model's decision-making process. Variables with higher absolute alphas have a more significant impact on the model's prediction. The analysis reveals that the ZHVI and ZORI are the most important variables, with alphas of 3.039253 and 2.677687, respectively. This suggests that these variables have a strong influence on the sale-to-list ratio. Other variables that have a relatively high impact on the model's prediction include Sale Price, List-to-Price Ratio, and New Construction Sale Price. On the other hand, Size Rank and Sale Price Change, have a negative impact on the model's prediction. It is important to

note, that the negative impact of Size Rank and Share of Listings with Price Cut variables on the model's prediction, as indicated by their negative SVM alphas, suggests that these features are inversely related to the target variable, which in this case is the sale-to-list ratio. For Size Rank, the negative alpha (-1.67285) indicates that as the Size Rank increases (representing smaller MSAs), there is a decrease in the predicted value of the target variable (e.g., sale-to-list ratio). This implies that smaller MSAs tend to have lower sale-to-list ratios compared to larger MSAs. This is also supported by Figure 3, in which larger MSAs exhibit higher sale-to-list ratio levels (New York has the lowest SR = 1, then Los Angeles, Chicago, Dallas, and finally Houston). Similarly, for the SLPC variable, a negative alpha suggests that as the share of listings with a price cut increases, there is a decrease in the predicted value of the target variable. This means that a higher proportion of listings with price cuts is associated with lower sale-to-list ratios. Therefore, the negative signs in these cases indicate an inverse relationship between the respective features and the target variable, consistent with the interpretation of negative coefficients in regression analysis.

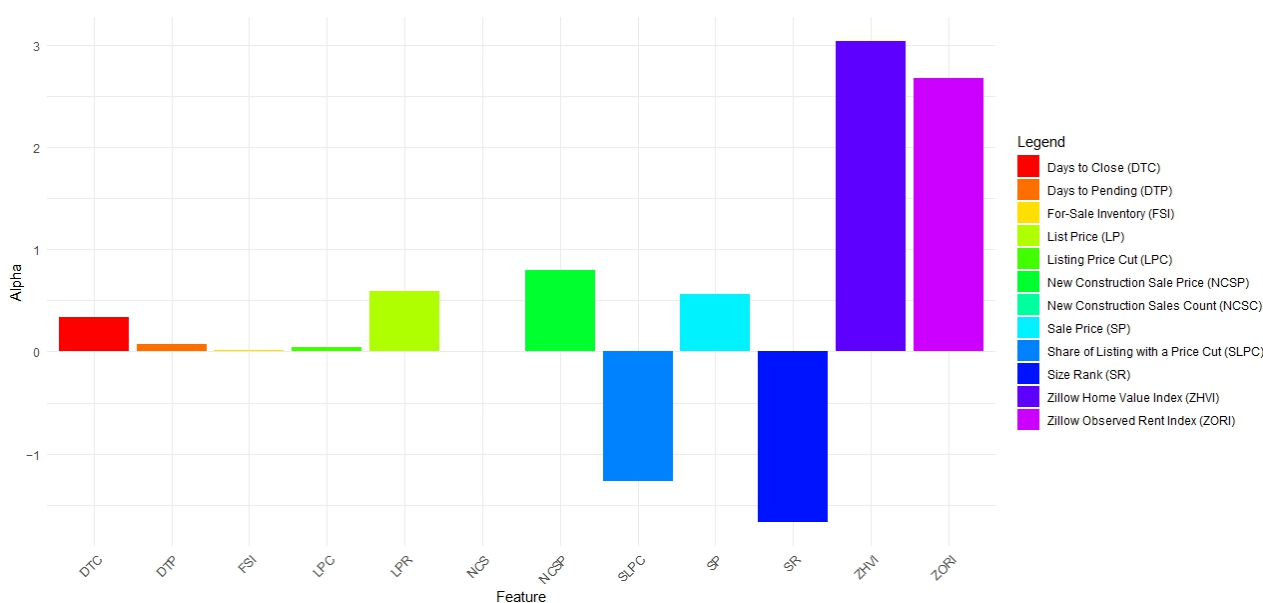


Figure 7. Alpha values (variable importance) for the SVM model (radial kernel); source: own elaboration in R.

These findings provide a unique lens through which to understand the interplay of factors influencing the sale-to-list ratio in the US real estate market.

Interestingly, contrary to conventional wisdom, our analysis uncovers an intriguing trend: properties with longer days to pending (DTP) and days to close (DTC) are associated with higher sale-to-list ratios (as can be seen in Table 7), challenging conventional assumptions about the relationship between transaction speed and market competitiveness in the real estate sector. This unexpected phenomenon may be due to various factors, such as increased demand for properties in certain neighborhoods or a perception of scarcity that leads buyers to act more decisively when faced with longer wait times. It also underscores the complex interplay between supply, demand and buyer behavior in shaping market dynamics.

Overall, the results of our study present a nuanced portrayal of the US real estate market. The amalgamation of accuracy scores, confusion matrix results, RFE insights, scaling effects, and SVM weights ensures a comprehensive evaluation of the SVM model's effectiveness. These findings lay a robust foundation for subsequent discussions, implications, and recommendations. The study leverages SVM classification and RFE to unravel nuanced relationships within the US real estate market. The identified key variables offer a foundation for more targeted analyses and informed decision-making in the real estate

domain. These results contribute to advancing our understanding of the intricate interplay of the factors influencing the dynamic and driving forces of the US real estate market, paving the way for more effective market predictions and strategic interventions.

The difference between the alphas approach and the RFE approach lies in the methods they use to identify important variables. The alphas approach, using SVM weights in the radial kernel, highlights the relative importance of variables based on their alphas, with higher absolute alphas indicating a more significant impact on the model's decision-making process. On the other hand, RFE is based on the idea of repeatedly constructing a model and choosing the best or worst performing feature, thus providing a greedy optimization for finding the best performing subset of features. The difference in these approaches can lead to variations in the identified important variables. Despite their differences, both approaches provide valuable insights into the interplay of factors influencing the sale-to-list ratio in the US real estate market. The alphas approach offers a quantitative measure of the relative importance of each variable, allowing for a clear ranking of their impact. On the other hand, RFE provides a systematic method for feature selection, identifying a subset of features that contribute most to the predictive ability of the model. By reconciling the results of both approaches, a more comprehensive understanding of the influential variables in predicting the sale-to-list ratio can be achieved.

In the case of the studied variable, the sale-to-list ratio, the contribution of both approaches can be reconciled by considering the commonalities in their findings. For instance, in the given scenario, both approaches point to the significance of the ZORI as an important variable. This convergence reinforces the importance of ZORI in predicting the sale-to-list ratio, providing a robust understanding of its influence. Additionally, the other variables identified by each approach can be further analyzed to understand their individual and collective impact on the sale-to-list ratio, thus enriching the overall analysis. Overall, while the alphas approach and RFE may yield different results due to their distinct methodologies, their combined insights offer a more holistic understanding of the variables influencing the sale-to-list ratio.

This study explores different approaches to understanding the key factors that influence sale-to-list ratio in the US real estate market. By examining various indicators such as Zillow indices, days to pending, days to close, listing price cut, and the share of listings with a price cut, the study uncovers previously undisclosed narratives within the market. The findings of this study can guide smarter decision-making for stakeholders interested in the US real estate market, equipping them with a strategic guide to navigate this intricate landscape. By combining advanced approaches and techniques with real-world insights, the study aims to make the real estate market more accessible and understandable for everyone involved. In other words, the study unveils the specific nuances of the US real estate market, emphasizing the influence of certain features on the sale-to-list ratio. The incorporation of SVM alphas enhances interpretability, providing practitioners and policymakers a more nuanced understanding.

It is important to note that the US real estate market is a dynamic ecosystem influenced by numerous factors. This study employs advanced machine learning techniques, including SVMs with radial kernel, to unravel the intricacies of this complex system. Our exploration goes beyond accuracy metrics, delving into the RFE method and, notably, SVM alphas, to provide a nuanced understanding of variable importance. Alpha values in SVM models quantify the impact of each variable on the target. Linear kernel SVMs result in straightforward alphas, directly reflecting variable importance. However, the radial kernel introduces complexity. Alphas in radial SVMs signify not only variable impact but also the influence of the data's non-linearity, capturing intricate patterns crucial for understanding the real estate landscape. RFE results identified key variables like days to pending, listing price cut, share of listings with a price cut, days to close, and ZORI. These align with the SVM alphas, providing a robust, corroborative understanding of their importance in predicting the sale-to-list ratio, and ultimately in predicting the driving forces of the US real estate market. These findings suggest potential strategies for market stakeholders. For

instance, focusing on Zillow indices, understanding the impact of days to pending and days to close, and recognizing the significance of listing price cuts, and the share of listings with a price cut can guide informed decision-making. Overall, the study emphasizes the importance of combining advanced machine learning techniques with domain-specific insights for a holistic understanding of real estate dynamics.

5. Discussion

In the realm of real estate markets, understanding seller behavior and pricing dynamics is crucial for predicting market forces and making informed decisions [62–64]. For example, Henriksson and Werlinder [62] demonstrate the importance of accurate real estate price prediction models, which rely on understanding the factors that influence seller pricing behavior. Their comparison of XGBoost and Random Forest models highlights how machine learning techniques can capture complex pricing dynamics in the housing market. Anenberg [63] examines how information frictions, such as sellers' and buyers' perceptions of market conditions, affect housing market dynamics and price formation. The study underscores the need to account for behavioral biases and asymmetric information to better predict market trends. Paraschiv and Chenavaz [64] specifically focus on the role of reference points in shaping seller pricing decisions. They show how sellers' reference points, shaped by past experiences and market conditions, can lead to loss aversion and influence listing prices. Understanding these behavioral factors is crucial for forecasting housing market outcomes. Several scientific studies delve into these aspects, shedding light on how sellers respond to market conditions, particularly during periods of housing busts and potential loss [65–67]. In this vein, the study by Zheng et al. [65] finds that sellers exhibit "speculative behavior" in the housing market, which can lead to boom and bust cycles. Specifically, the authors show that during a housing boom, sellers become more reluctant to lower their asking prices, even as demand starts to decline. This is because sellers are influenced by the recent high prices and are unwilling to accept lower offers, even if it means their homes sit on the market longer than they would like. Agnello and Schuknecht [66] analyzed the determinants and implications of booms and busts in housing markets. Their research indicates that sellers are reluctant to lower prices during housing busts, contributing to the persistence of high prices and reduced market activity. Agnello et al. [67] further explored the dynamics of booms, busts, and normal times in the housing market. They found that sellers' loss aversion behavior, where they are unwilling to sell at a loss, can prolong housing market downturns.

The exploration into the dynamics of the US real estate market has involved meticulous methodological choices, rigorous data analysis, and a pursuit of actionable insights. When discussing the findings, it is crucial to critically evaluate the background, methodologies, and results to discern the strengths and the areas for potential refinement in the approach. The backdrop against which this study unfolds is grounded in the fundamental importance of the sale-to-list ratio in the US real estate market, serving as a barometer for market conditions and reflecting the interplay between buyer and seller dynamics [9]. The fundamental importance of the sale-to-list ratio in the real estate market is well-established [9–13]. Several studies have highlighted the significance of the sale-to-list ratio in understanding market dynamics and informing decision-making processes. For example, Zhang et al. [9] focused on the relationship between sale-to-list ratio and for-sale inventory, emphasizing its significance in the real estate market. Vaidynathan et al. [68] investigated the effects of economic factors on the median list price and median selling price in the US housing market, indirectly highlighting the importance of the sale-to-list ratio in understanding market dynamics. Bich et al. [34] explored the dynamic effects of listing price strategies on the probability of selling a house, which is related to the interplay between list price and sale price, contributing to the understanding of the sale-to-list ratio's importance in the real estate market. These studies support the fundamental importance of sale-to-list ratio in the real estate and housing markets, emphasizing its role as a barometer for market conditions and reflecting the dynamics between buyers and sellers.

The decision to employ Support Vector Machines for classification purposes stems from its inherent capacity to handle complex, nonlinear relationships in the data [69–72]. Unlike traditional regression models, a SVM provides a powerful tool for predicting sale-to-list ratios, capturing intricate patterns that might elude linear approaches. The Recursive Feature Elimination method, integrated into the SVM framework, adds an additional layer of insight by systematically identifying the most influential features [51,52,73–75]. This combination allows for a nuanced understanding of the US real estate market dynamics. However, the reliance on SVMs does not come without considerations. While a SVM excels in capturing intricate relationships, its interpretability may pose challenges, as highlighted by the opaque nature of the radial kernel weights, known as alphas [76–79]. The issue of interpretability has been addressed in the literature. For instance, Hakkoum et al. [76] emphasize the importance of interpretability in medical applications, highlighting its role in enhancing trust in machine learning models. Abuali et al. [77] discuss the use of SVMs in intrusion detection systems, noting the need for interpretability to improve the reliability of these models. Samuel et al. [78] propose the use of SVMs in explainable AI (XAI) models to improve interpretability and accuracy, acknowledging that SVMs are more interpretable than complex neural networks but still require augmentation with more descriptive explanations provided by medical experts. Valentin et al. [79] examine the difference between relevance and reliance on predictor variables in multivariate models, emphasizing that SVMs with an RBF kernel function were used in their study, achieving a cross-validated classification accuracy of 48.21%. The interpretability concern underscores the importance of supplementary analyses, such as the RFE, to shed light on variable importance. Despite this challenge, the trade-off in predictive power and complexity is justifiable given the complex nature of real estate dynamics. The lessons learned from medicine and other sciences can be applied to the real estate context, where the inherent complexity of the market requires a nuanced approach to model development and interpretation.

The array of results presented in this study opens a window into the nuanced interplay of variables influencing the sale-to-list ratio. The high accuracy score, approaching 85%, signifies the effectiveness of the SVM model in predicting this ratio. The confusion matrix, with its precision, recall, and F1 scores, adds granularity to the understanding, pinpointing areas of model strength and potential improvement [80–82]. Precision and recall are crucial for information retrieval and other applications, where false positives and false negatives have different costs [80–82]. The F1 score, which is the harmonic mean of precision and recall, provides a balanced view of the model's performance [80–82]. The confusion matrix, precision, recall, and F1 scores are essential for evaluating classifiers and their effectiveness, especially in multi-class classification scenarios, often involving imbalanced datasets [82]. The 85% accuracy indicates that the model is able to correctly classify sale-to-list ratios into their respective categories about 85% of the time. In simpler terms, if we use this model to predict whether a property's sale price will accurately match its listing price, it will be correct about 85 out of 100 times. For real estate stakeholders such as buyers, sellers, or investors, this level of accuracy can have significant practical implications. For example, a seller could use this model to determine an appropriate listing price based on the predicted sale-to-list ratio, helping them make more informed pricing decisions and potentially maximize their returns. Similarly, a buyer could use the model's predictions to evaluate the competitiveness of a property's listing price and negotiate accordingly. Overall, the model's accuracy provides stakeholders with valuable insights into market trends and helps guide their decision-making processes in the real estate market.

The Recursive Feature Elimination results highlight specific variables—days to pending (DTP), listing price cut (LPC), share of listings with a price cut (SLPC), days to close (DTC), and ZORI—as crucial drivers of the sale-to-list ratio. These findings align with industry expectations and provide a robust foundation for targeted interventions and strategic decision-making. Prior research has demonstrated the utility of RFE in identifying key features for predicting house prices in the United States. For instance, Wu [83] employed RFE as one of several feature selection methods to forecast housing prices in King County,

USA, and found that it aids in selecting important features and preventing over-fitting the model with an excessive number of features. Similarly, the study by Yang et al. [84] utilized RFE as part of a feature selection process to predict house prices in Ames, Iowa, and concluded that it is effective in identifying the most crucial features for price prediction.

While the study provides valuable insights, it is essential to acknowledge the inherent challenges of navigating the complexities of the real estate market. The reliance on machine learning models, particularly SVM, necessitates a delicate balance between predictive power and interpretability [85]. Interpretable machine learning involves extracting relevant knowledge from a model concerning relationships in the data or learned by the model. It emphasizes the degree to which a human can understand the cause of a decision, aiding in providing actionable insights [86,87]. The challenge lies in ensuring that the models not only forecast accurately but also offer actionable insights that resonate with industry stakeholders [85].

The interpretation of the SVM radial basis kernel alpha coefficients, acting as weighting parameters, presents a significant challenge due to their role in capturing variable importance yet potentially contributing to the “black box” nature of the model. This impediment underscores the necessity of clear and effective communication of results, as stakeholders demand both precise forecasts and an understandable narrative that supports informed decision-making. Notably, the recent literature has highlighted the relevance of carefully selecting appropriate kernel functions and suitable training datasets within SVM models, thereby accentuating the need to address interpretability issues when communicating outcomes to stakeholders. For instance, Yekkehkhany et al.’s [88] work, comparing linear, polynomial, and RBF kernels for SVM-based classification, sheds light on the importance of selecting the most fitting kernel function according to specific dataset features. Additionally, Nalepa and Kawulok’s [89] study addresses various challenges associated with SVMs, such as parameter selection and working with low-quality data, touching upon the significance of comprehending which vectors are chosen as support vectors, thus enhancing the interpretability of SVM decisions.

In accordance with the obtained findings, several strategic suggestions arise, drawing upon both the empirical outcomes and extant industrial perspectives. Firstly, the prominence of variable significance emphasizes the necessity for ongoing investments in domain expertise to ensure the relevance and applicability of predictive models in real-world decision-making processes. Collaborative efforts among industry professionals and data science experts are crucial to guarantee alignment between selected variables and market intricacies alongside emerging trends. Secondly, the potential for exploring hybrid modeling approaches that combine the predictive power of SVMs with the interpretability of regression-based models is identified. It relies on the study’s framework that leverages the strengths of both techniques to accurately predict and provide insights into the key drivers of the U.S. real estate market’s sale-to-list ratio. Such methodologies might serve as a conduit between precision and clarity, thereby addressing the divergent demands of various stakeholders. Thirdly, the scalability of the proposed model for its application within diverse real estate markets necessitates further investigation. Although the current study concentrates on the top 100 MSAs, extending the model’s reach to encompass regional and local markets would offer valuable insights for localized decision-making processes. Overall, the examination of US real estate market dynamics accentuates the imperativeness of striking a balance between predictive capability and comprehensibility, as demonstrated by the integration of robust predictive models (SVMs) and interpretable feature selection (RFE) to deliver accurate forecasts and actionable insights into the key drivers of real estate market trends. By critically evaluating the techniques employed, appreciating the nuances that arise from the results, and providing practical guidance, this endeavor aims to provide a sophisticated understanding of the multifaceted real estate market. This study aims to guide the evolution of real estate prognostication, making it more nuanced and accurate.

The results of our study provide a rich understanding of the dynamics of the U.S. real estate market and shed light on key variables that significantly impact the sale-to-

list ratio, a key metric that indicates the health and competitiveness of the US real estate market and whether a market favors buyers or sellers [5–7]. Through a combination of advanced machine learning methods, including SVM with radial kernel and Recursive Feature Elimination, the study provides insights that can inform both industry stakeholders and policymakers. It is important to study the implications of our findings, particularly focusing on the five variables identified through the RFE approach—Days to Pending, Listing Price Cut, Share of Listings with Price Cut, Days to Close, and ZORI. When it comes to ‘Days to Pending’ and ‘Days to Close’, these temporal indicators play a crucial role in understanding the pace and activity levels within the US real estate market. The shorter the days to pending and days to close, the more active the market tends to be, indicating higher demand and faster turnover of properties. The study by Truong et al. [19] corroborates this notion, highlighting the significance of time-related variables in predicting housing prices accurately. Additionally, Henriksson and Werlinder [62] explore the performance of machine learning models on housing price data, and they mention the effects of data variation, which can include time-related factors, on the accuracy of predictions. Moreover, Carrillo’s [5] study on seller and buyer bargaining power underscores the importance of understanding the temporal dynamics in real estate transactions. Shorter days to pending and days to close suggest a seller’s market, where properties are in high demand and buyers may face increased competition, potentially leading to higher sale-to-list ratios. With this study we examine the relationship between property listing metrics and market dynamics, focusing specifically on the duration of “days to pending” and “days to close”. Our findings suggest that shorter intervals for both metrics are associated with a seller’s market characterized by heightened demand and increased competition among prospective buyers. To contextualize our observations, we draw on data from a comprehensive study conducted by Zillow [90] that sheds light on the factors that contribute to failed transactions involving properties listed as pending. The Zillow report reveals several common causes of unsuccessful closings, including financial complications, inadequate inspections, and valuation inconsistencies. These challenges underscore the need for efficient navigation in a highly competitive environment, where shorter timelines from pending status to closing mean greater pressure on buyers and sellers alike to address such obstacles promptly [90]. As a result, our findings support the notion that reduced days to pending and days to close are consistent with a seller’s market characterized by increased demand and competitiveness, ultimately resulting in a higher sale-to-list ratio. Overall, the study examines the interplay between real estate listings and market conditions and supports the assumption that shorter “days to pending” and “days to close” periods indicate a seller’s market characterized by robust demand and increased competition among buyers, resulting in higher sale-to-list ratios. These conclusions are supported by empirical evidence from the Zillow study [90], which illustrates the complexity of navigating the intricacies of a dynamic housing market.

As for the ‘Listing Price Cut’ and ‘Share of Listings with Price Cut’, these variables reflect price dynamics within the market and can serve as indicators of market competitiveness and seller flexibility. A high listing price cut or a significant share of listings with price cuts may indicate an oversaturated market or sellers’ willingness to negotiate, potentially leading to lower sale-to-list ratios. This aligns with the findings of Miller and Sklarz [6], who emphasize the importance of pricing indicators in short-term price forecasting and market condition assessment. Understanding the prevalence and impact of price cuts can provide valuable insights for both buyers and sellers, guiding pricing strategies and negotiation tactics. Here, of particular note, is the study by Keys and Mulder [91], which sheds light on the complexity of the housing market dynamics and the consequences of failing to respond appropriately to changing market conditions. Keys and Mulder [91] examine the relationship between exposure to sea level rise and changes in the housing and mortgage markets over the 2001–2020 period, focusing specifically on coastal Florida. The authors find that while transaction volumes began to decline in 2013, prices did not follow suit until several years later. This suggests that sellers’ pricing strategies did not

initially adjust for the risks associated with sea level rise, leading to a mismatch between buyers' expectations and sellers' valuations. While their study does not explicitly address price cuts in the traditional sense, it provides insight into the dynamic interplay between environmental factors, housing market behavior, and pricing strategies. It offers a nuanced view of how external forces influence real estate markets and highlights the potential consequences when sellers fail to adequately consider evolving market conditions. This knowledge may indirectly contribute to our understanding of the benefits of pursuing price cuts within the US housing market.

In turn, the ZORI offers insights into rental market conditions, which can influence housing demand and pricing dynamics. The study by Ghosalkar and Dhage [23] underscores the interconnectedness between rental and housing markets, highlighting the potential impact of rental trends on housing prices. A rising ZORI may indicate increased demand for rental properties, potentially spurring demand for homeownership and influencing the sale-to-list ratio. Conversely, a declining ZORI may suggest shifts in housing preferences or affordability constraints, impacting housing market dynamics accordingly. The factors influencing the rental housing market, including demand for rental properties, are multifaceted and can have ripple effects on other segments of the housing market such as homeownership. Several key factors affecting rental demand and prices include supply and demand dynamics [91], property specific attributes [92], economic factors [93], and market supply and demand [94].

By contextualizing these variables within the broader literature on the US housing market, we can better understand their implications for market participants. The interplay between temporal indicators, pricing dynamics, and rental market conditions underscores the complexity of the real estate landscape and the importance of considering multiple factors in real estate market analysis. These are complex scientific problems to assess and as such they also require more sophisticated approaches to their analyses. Recent scientific evidence shows that, to solve such dynamic problems, researchers more often reach out to machine learning method or other complex approaches. In this vein, Grybauskas et al. [95] examined the impact of the COVID-19 pandemic on the real estate market in Lithuania using big data and machine learning techniques. Grybauskas et al. [95] used an XGBoost model for predictive analytics on the real estate market during the COVID-19 pandemic. Grybauskas et al.'s [95] study found that the "time on the market" (TOM) variable (TOM refers to the number of days between the listing date and the off-the-market date) was the most dominant and consistent predictor of apartment prices, exhibiting an inverse U-shaped relationship. This suggests that both very short and very long TOM values can signal emerging problems in the market that could lead to recessions or overheating. The authors recommend that governments and investors closely monitor TOM values as they provide useful real-time information about market conditions.

Also, Gabrovski and Ortego-Martí (2018) developed a search-and-matching model to analyze housing market dynamics, specifically focusing on the role of search frictions. Their model examines the relationship between the stock of houses for sale (and equivalent of for-sale inventory), house prices, and the time it takes to sell a house (time on the market). Importantly, their study shows that the time on the market is closely linked to housing market tightness, which is defined as the ratio of buyers to sellers. Tighter housing markets (more buyers relative to sellers) are associated with shorter time on the market. In this regard, the time on the market variable can be viewed as an alternative to metrics like Zillow's 'days to pending' (DTP) or 'days to close' (DTC). This is because the time on the market captures the overall duration from listing to sale, rather than just the time to get a pending offer or close the transaction. The model is able to generate procyclical vacancies and sales, as well as countercyclical time on the market, which aligns with empirical observations of housing market behavior. This might explain to some extent the seasonality (cyclicality) that can be viewed in Figures 2 and 3.

Overall, this study uses advanced analytical techniques, specifically SMV and RFE, to examine the complex relationships that govern U.S. real estate property values as indicated

by the sale-to-list ratio. Our findings contribute to the advancement of knowledge in the field of real estate research and provide valuable insights into the U.S. housing market. In addition, our study provides actionable advice for stakeholders navigating the intricacies of real estate investment and decision-making. More specifically, it provides tangible guidance for real estate stakeholders, backed by empirical evidence and advanced analytical techniques. For example, based on the findings from this study, real estate agents and developers can adjust their pricing strategies by considering variables such as days to pending (DTP) and listing price cuts (LPC). For example, contrary to conventional wisdom, our analysis uncovers an intriguing trend: properties with longer days pending and days to close are associated with higher sale-to-list ratios, challenging conventional assumptions about the relationship between transaction speed and market competitiveness in the real estate sector. This unexpected phenomenon may be due to various factors, such as increased demand for properties in certain neighborhoods or a perception of scarcity that leads buyers to act more decisively when faced with longer wait times. It also underscores the complex interplay between supply, demand and buyer behavior in shaping market dynamics. In addition, the research highlights the importance of rental market conditions, as captured by the ZORI, in influencing sale-to-list ratios. Armed with this knowledge, investors and landlords can make informed decisions regarding rental property acquisitions, taking into account the rental market trends to maximize returns. In addition, our study underscores the importance of collaboration between industry professionals and data science experts to ensure alignment between selected variables and emerging market trends. By incorporating these actionable insights into their decision-making processes, stakeholders can adapt to evolving market conditions and capitalize on opportunities in the real estate sector. As the market undergoes continuous evolution, the insights from our study can serve as a valuable resource for adapting to changing conditions and capitalizing on opportunities in the real estate sector.

Furthermore, the comparison between SVM alphas and RFE highlights the complementary nature of these approaches in identifying influential variables. While SVM alphas offer a quantitative measure of variable importance, RFE provides a systematic method for feature selection. By reconciling the results of both approaches, we gain a more comprehensive understanding of the factors driving the sale-to-list ratio and, by extension, the broader dynamics of the US real estate market.

Last but not least, while this study offers valuable insights into the predictive power of various real estate metrics on the sale-to-list ratio in the US housing market, it is not without limitations. Firstly, the reliance on a single dataset from Zillow Econ may introduce biases or limitations inherent in that dataset, potentially affecting the generalizability of the findings. Additionally, the study's focus on major regions by MSA may overlook nuances or variations in real estate dynamics at smaller geographical scales, limiting the applicability of the results to broader contexts. Furthermore, while the SVM classification algorithm and RFE method are powerful analytical tools, they may not capture all relevant factors influencing the sale-to-list ratio, leaving room for unexplored variables or interactions. Moreover, the interpretation of SVM weights (alphas) may pose challenges due to the complexity of the radial kernel, potentially limiting the depth of insight into variable importance. Lastly, the study's retrospective design limits its ability to establish causal relationships between the identified real estate metrics and the sale-to-list ratio, warranting further longitudinal or experimental research to confirm the observed associations.

6. Conclusions

This study delves into the workings of the US real estate market, specifically looking at the sale-to-list ratio, a measure that reflects the vitality and competitiveness of the housing sector. The aim of the analysis was to gain an insight into how various factors impact the sale-to-list ratio given the evolving market dynamics and growing complexity of data sources. By utilizing machine learning methods like radial kernel Support Vector Machines,

Recursive Feature Elimination, and SVM alphas, this study offers an examination of the key factors influencing the sale-to-list ratio.

The research methodology employed a systematic approach to data analysis, starting with the collection of a robust dataset spanning 100 major US regions over a period of 70 months. The choice of SVM classification, known for its ability to capture non-linear relationships within data, was influenced by the complexity inherent in real estate dynamics. RFE served as a crucial component in distilling the extensive set of predictor variables, systematically pruning less informative features to refine the model's predictive power. Additionally, using SVM alphas helped us gain insights into the importance of each variable in predicting the sale-to-list ratio and making the model easier to understand.

The study uncovered details about what influences the sale-to-list ratio in the US real estate market. The SVM model showed an 85% accuracy rate and a Kappa score of 77.4%, demonstrating its effectiveness in forecasting the sale-to-list ratio. Confusion matrix analysis further elucidated the model's performance, showcasing high sensitivity and specificity across various market segments. Notably, the top five variables identified through the RFE method—days to pending, listing price cut, share of listings with a price cut, days to close, and ZORI—emerged as critical predictors, offering valuable insights into market dynamics.

The significance of these discoveries extends beyond academic discussions, offering valuable insights for real estate researchers, policymakers, and investors. Real estate experts can use the identified factors to improve pricing strategies and successfully adapt to changes in the market. Policymakers can gain insights into the interconnected elements that impact housing markets, helping them create policies to tackle market issues. Investors, equipped with an understanding of factors, can make well-informed choices to handle risks and maximize profits. From an academic perspective, this study enhances the comprehension of the intricate factors influencing US real estate values. By integrating advanced machine learning methods with domain-specific knowledge, the research presents a comprehensive approach to analyzing real estate valuation dynamics. The utilization of SVM classification and RFE forms a robust framework for exploring complex relationships within real estate data, paving the way for future research endeavors in this field.

In summary, this study represents a significant addition to the real estate science literature, offering both theoretical insights and practical applications. By uncovering the determinants of the sale-to-list ratio and providing actionable recommendations, this paper aims to guide real estate decision-making processes and contribute to ongoing discussions on real estate market dynamics.

Author Contributions: Conceptualization, J.S.; methodology, J.S. and D.M.; validation, J.S.; investigation, J.S. and D.M.; resources, J.S.; data curation, D.M.; writing—original draft preparation, J.S. and D.M.; writing—review and editing J.S. and D.M.; visualization, D.M.; supervision, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data was collected from [<https://www.zillow.com/research/data>, accessed on 5 February 2024] and transformed into a panel data frame. We confirm that the data, models and methodology used in the research are proprietary, and the derived data supporting the findings of this study are available from the first author on request.

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

References

1. Sobieraj, J.; Metelski, D. Private Renting vs. Mortgage Home Buying: Case of British Housing Market—A Bayesian Network and Directed Acyclic Graphs Approach. *Buildings* **2022**, *12*, 189. [[CrossRef](#)]
2. Bryx, M.; Sobieraj, J.; Metelski, D.; Rudzka, I. Buying vs. Renting a Home in View of Young Adults in Poland. *Land* **2021**, *10*, 1183. [[CrossRef](#)]
3. Sobieraj, J.; Bryx, M.; Metelski, D. Preferences of Young Polish Renters: Findings from the Mediation Analysis. *Buildings* **2023**, *13*, 920. [[CrossRef](#)]

4. Raza, S.A.; Guesmi, K. Guest editorial: Predictability of housing prices in the times of crises: New trends, methodologies, and techniques. *Int. J. Hous. Mark. Anal.* **2024**, *17*, 1–7. [[CrossRef](#)]
5. Carrillo, P.E. To sell or not to sell: Measuring the heat of the housing market. *Real Estate Econ.* **2013**, *41*, 310–346. [[CrossRef](#)]
6. Miller, N.; Sklarz, M. Integrating real estate market conditions into home price forecasting systems. *J. Hous. Res.* **2012**, *21*, 183–213. [[CrossRef](#)]
7. Vinsand, E.; Sjong, H.H. Measuring The Heat of Oslo’s Housing Market: A Composite Indicator to Improve the Informational Efficiency in the Residential Real Estate Market. Master’s Thesis, Norwegian School of Economics, Bergen, Norway, 2021.
8. Anenberg, E.; Ringo, D. Volatility in Home Sales and Prices: Supply or Demand? *J. Urban Econ.* **2024**, *139*, 103610. [[CrossRef](#)]
9. Zhang, X.; Wang, Y.; Yang, E.; Xu, S.; Yu, Y. Sale to list ratio, for-sale inventory, sale count, and housing value. *International J. Hous. Mark. Anal.* **2023**, *1*, 1–15. [[CrossRef](#)]
10. Damen, S. The quality-adjusted price evolution of houses and apartments in Flanders from 2005q1 until 2021q2. *SSRN Electron. J.* **2021**, 1–6. [[CrossRef](#)]
11. Hattapoglu, M.; Hoxha, I. Hot and cold seasons in Texas housing markets. *Int. J. Hous. Mark. Anal.* **2021**, *14*, 317–332. [[CrossRef](#)]
12. Schmidbauer, E. Time on the Market and List Prices in “Hot” Real Estate Markets. *SSRN Electron. J.* **2023**, 4331752, 1–28. [[CrossRef](#)]
13. Hoxha, I.; Yilmaz, S. Liquidity and pricing trends in the housing market in Pennsylvania. *Pa. Econ. Rev.* **2020**, *27*, 1–16.
14. Huang, Y.; Yip, T.L.; Liang, C. Risk Perception and Property Value: Evidence from Tianjin Port Explosion. *Sustainability* **2020**, *12*, 1169. [[CrossRef](#)]
15. Engerstam, S.; Warsame, A.; Wilhelmsson, M. Long-term dynamics of new residential supply: A case study of the apartment segment in Sweden. *Buildings* **2022**, *12*, 970. [[CrossRef](#)]
16. Riccioli, F.; Fratini, R.; Boncinelli, F. The impacts in real estate of landscape values: Evidence from Tuscany (Italy). *Sustainability* **2021**, *13*, 2236. [[CrossRef](#)]
17. Szczepańska, A.; Gościewski, D.; Gerus-Gościewska, M. A GRID-based spatial interpolation method as a tool supporting real estate market analyses. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 39. [[CrossRef](#)]
18. Antipov, E.A.; Pokryshevskaya, E.B. Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Syst. Appl.* **2012**, *39*, 1772–1778. [[CrossRef](#)]
19. Truong, Q.; Nguyen, M.; Dang, H.; Mei, B. Housing price prediction via improved machine learning techniques. *Procedia Comput. Sci.* **2020**, *174*, 433–442. [[CrossRef](#)]
20. Geerts, M.; De Weerd, J. A Survey of Methods and Input Data Types for House Price Prediction. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 200. [[CrossRef](#)]
21. Liu, G. Research on prediction and analysis of real estate market based on the multiple linear regression model. *Sci. Program.* **2022**, *2022*, 5750354. [[CrossRef](#)]
22. Manjula, R.; Jain, S.; Srivastava, S.; Kher, P.R. Real estate value prediction using multivariate regression models. *IOP Conf. Ser. Mater. Sci. Eng.* **2017**, *263*, 042098. [[CrossRef](#)]
23. Ghosalkar, N.N.; Dhage, S.N. Real estate value prediction using linear regression. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBEA), Pune, India, 16–18 August 2018.
24. Michele, L.; Andrea, L.; Pangallo, M. What do online listings tell us about the housing market? *Int. J. Cent. Bank.* **2022**, *1*, 1–53.
25. Ho, W.K.; Tang, B.S.; Wong, S.W. Predicting property prices with machine learning algorithms. *J. Prop. Res.* **2021**, *38*, 48–70. [[CrossRef](#)]
26. Alzain, E.; Alshebami, A.S.; Aldhyani, T.H.H.; Alsubari, S.N. Application of Artificial Intelligence for Predicting Real Estate Prices: The Case of Saudi Arabia. *Electronics* **2022**, *11*, 3448. [[CrossRef](#)]
27. Yilmaz, O.; Talavera, O.; Jia, J.Y. Rental market liquidity, seasonality, and distance to universities. *Int. J. Econ. Bus.* **2022**, *29*, 223–239. [[CrossRef](#)]
28. Anenberg, E.; Ringo, D. *Housing Market Tightness During COVID-19: Increased Demand or Reduced Supply?* Board of Governors of the Federal Reserve System: Washington, DC, USA, 2021.
29. Ngai, L.R.; Sheedy, K.D. The decision to move house and aggregate housing-market dynamics. *J. Eur. Econ. Assoc.* **2020**, *18*, 2487–2531. [[CrossRef](#)]
30. Gabrovski, M.; Ortego-Martí, V. *Housing Market Dynamics with Search Frictions (No. 201804)*; University of California: Riverside, CA, USA, 2018; pp. 1–27.
31. Ngai, L.R.; Tenreiro, S. Hot and cold seasons in the housing market. *Am. Econ. Rev.* **2014**, *104*, 3991–4026. [[CrossRef](#)]
32. Guren, A. *The Causes and Consequences of House Price Momentum*; Harvard University: Cambridge, MA, USA, 2014; pp. 1–106.
33. Diaz, A.; Jerez, B. House prices, sales, and time on the market: A search-theoretic framework. *Int. Econ. Rev.* **2013**, *54*, 837–872. [[CrossRef](#)]
34. Bich, H.N.T.; Trong, H.N.; Thanh, H.T. The role of listing price strategies on the probability of selling a house: Evidence from Vietnam. *Real Estate Manag. Valuat.* **2020**, *28*, 63–75. [[CrossRef](#)]
35. Leamer, E.E. *Housing Is the Business Cycle (Working Paper 13428)*; National Bureau of Economic Research: Cambridge, MA, USA, 2007.
36. Gilbukh, S.; Goldsmith-Pinkham, P. *Heterogeneous Real Estate Agents and the Housing Cycle (No. w31683)*; National Bureau of Economic Research: Cambridge, MA, USA, 2023.

37. Sklarz, M. Days on Market: The Unbearable Tightness of Inventory. Available online: <https://www.blackknightinc.com/blog/the-unbearable-tightness-of-inventory/> (accessed on 10 April 2024).
38. Schölkopf, B.; Smola, A.J.; Bach, F. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.
39. Shin, K.S.; Lee, T.S.; Kim, H.J. An application of Support Vector Machines in bankruptcy prediction model. *Expert Syst. Appl.* **2005**, *28*, 127–135. [[CrossRef](#)]
40. Osuna, E.; Freund, R.; Girosit, F. Training Support Vector Machines: An Application to Face Detection. Computer Vision and Pattern Recognition. In Proceedings of the IEEE Computer Society Conference, Los Alamitos, CA, USA, 17–19 June 1997.
41. Karatzoglou, A.; Meyer, D.; Hornik, K. Support Vector Machines in R. *J. Stat. Softw.* **2006**, *15*, 1–28. [[CrossRef](#)]
42. Platt, J. *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*; Technical Report MSR-TR-98-14; Microsoft Research: Redmond, WA, USA, 1998; pp. 1–21.
43. Platt, J. Using analytic QP and sparseness to speed training of Support Vector Machines. *Adv. Neural Inf. Process. Syst.* **1999**, *11*, 557–563.
44. Vishwanathan, S.V.; Smola, A.J.; Murty, M.N. Simple SVM. In Proceedings of the Twentieth International Conference on International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003.
45. Vapnik, V. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998.
46. Chapelle, O.; Vapnik, V. Model selection for Support Vector Machines. In *Advances in Neural Information Processing Systems*; Solla, S., Leen, T., Müller, K., Eds.; NIPS: Denver, CO, USA, 1999; pp. 230–236.
47. Bousquet, O.; Boucheron, S.; Lugosi, G. Introduction to Statistical Learning Theory. In *Advanced Lectures on Machine Learning*; Bousquet, O., von Luxburg, U., Rätsch, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 169–207.
48. Gunn, S.R. Support Vector Machines for classification and regression. *ISIS Tech. Rep.* **1998**, *14*, 5–16.
49. Salcedo-Sanz, S.; Rojo-Álvarez, J.L.; Martínez-Ramón, M.; Camps-Valls, G. Support Vector Machines in engineering: An overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2014**, *4*, 234–267. [[CrossRef](#)]
50. Brownlee, J. Recursive Feature Elimination (RFE) for Feature Selection in Python. Available online: <https://machinelearningmastery.com/rfe-feature-selection-in-python/> (accessed on 10 December 2023).
51. Huang, M.L.; Hung, Y.H.; Lee, W.M.; Li, R.K.; Jiang, B.R. SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. *Sci. World J.* **2014**, *1*, 795624. [[CrossRef](#)]
52. Sanz, H.; Valim, C.; Vegas, E.; Oller, J.M.; Reverter, F. SVM-RFE: Selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinform.* **2018**, *19*, 432. [[CrossRef](#)]
53. Fu, R.; Jin, G.Z.; Liu, M. Does Human-Algorithm Feedback Loop Lead to Error Propagation? Evidence from Zillow’s Zestimate; National Bureau of Economic Research: Cambridge, MA, USA, 2023; pp. 1–44.
54. Yörük, B.K. Early effects of COVID-19 pandemic-related state policies on housing market activity in the United States. *J. Hous. Econ.* **2022**, *57*, 101857. [[CrossRef](#)]
55. Knight, J.R. Listing price, time on market, and ultimate selling price: Causes and effects of listing price changes. *Real Estate Econ.* **2002**, *30*, 213–237. [[CrossRef](#)]
56. Harker, P.T. *The Fed and the Economy: Where We’re Going, Where We’ve Been (No. 95648)*; La Salle University: Philadelphia, PA, USA, 2023; pp. 1–10.
57. Zillow. What Is Zillow’s Buyer-Seller Index, and How Is It Computed? Available online: <https://www.zillow.com/research/understanding-the-zillow-buyer-seller-index-2883/> (accessed on 10 April 2024).
58. Pariser, I. *The Effect of School Quality on Prices versus Rents*; University of California: Berkeley, CA, USA, 2019; pp. 1–48.
59. Kotova, N.; Zhang, A.L. Liquidity in residential real estate markets. In Proceedings of the ASSA 2021 Annual Meeting, Held Virtually, 3–5 January 2021.
60. Maxwell, T.; Segal, T. Sale-to-List Ratio: Why It Matters. Available online: <https://www.bankrate.com/real-estate/sale-to-list-ratio/> (accessed on 20 November 2023).
61. StreetEasy Team. What Is the Sale-to-List Price Ratio, and Why Does It Matter? Available online: <https://streeteasy.com/blog/sale-to-list-price-ratio/> (accessed on 20 November 2023).
62. Henriksson, E.; Werlinder, K. *Housing Price Prediction over Countrywide Data: A Comparison of XGBoost and Random Forest Regressor Models*; KTH, School of Electrical Engineering and Computer Science (EECS): Stockholm, Sweden, 2021.
63. Anenberg, E. Information frictions and housing market dynamics. *Int. Econ. Rev.* **2016**, *57*, 1449–1479. [[CrossRef](#)]
64. Paraschiv, C.; Chenavaz, R. Sellers’ and buyers’ reference point dynamics in the housing market. *Hous. Stud.* **2011**, *26*, 329–352. [[CrossRef](#)]
65. Zheng, M.; Wang, H.; Wang, C.; Wang, S. Speculative behavior in a housing market: Boom and bust. *Econ. Model.* **2017**, *61*, 50–64. [[CrossRef](#)]
66. Agnello, L.; Schuknecht, L. Booms and busts in housing markets: Determinants and implications. *J. Hous. Econ.* **2011**, *20*, 171–190. [[CrossRef](#)]
67. Agnello, L.; Castro, V.; Sousa, R.M. Booms, busts, and normal times in the housing market. *J. Bus. Econ. Stat.* **2015**, *33*, 25–45. [[CrossRef](#)]
68. Vaidynathan, D.; Kayal, P.; Maiti, M. Effects of economic factors on median list and selling prices in the US housing market. *Data Sci. Manag.* **2023**, *6*, 199–207. [[CrossRef](#)]

69. Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **2020**, *408*, 189–215. [[CrossRef](#)]
70. Durgesh, K.S.; Lekha, B. Data classification using support vector machine. *J. Theor. Appl. Inf. Technol.* **2010**, *12*, 1–7.
71. Zhang, Y. Support vector machine classification algorithm and its application. In Proceedings of the Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, 14–16 September 2012.
72. Anthony, G.; Greg, H.; Tshilidzi, M. Classification of images using Support Vector Machines. *arXiv* **2007**, arXiv:0709.3967.
73. Han, Y.; Huang, L.; Zhou, F. A dynamic Recursive Feature Elimination framework (dRFE) to further refine a set of OMIC biomarkers. *Bioinformatics* **2021**, *37*, 2183–2189. [[CrossRef](#)]
74. Guo, Y.; Zhang, Z.; Tang, F. Feature selection with kernelized multi-class support vector machine. *Pattern Recognit.* **2021**, *117*, 107988. [[CrossRef](#)]
75. Lin, X.; Yang, F.; Zhou, L.; Yin, P.; Kong, H.; Xing, W.; Xu, G. A support vector machine-Recursive Feature Elimination feature selection method based on artificial contrast variables and mutual information. *J. Chromatogr. B* **2012**, *910*, 149–155. [[CrossRef](#)]
76. Hakkoum, H.; Abnane, I.; Idri, A. Interpretability in the medical field: A systematic mapping and review study. *Appl. Soft Comput.* **2022**, *117*, 108391. [[CrossRef](#)]
77. Abuali, K.M.; Nissirat, L.; Al-Samawi, A. Advancing Network Security with AI: SVM-Based Deep Learning for Intrusion Detection. *Sensors* **2023**, *23*, 8959. [[CrossRef](#)] [[PubMed](#)]
78. Samuel, S.S.; Abdullah, N.N.; Raj, A. Interpretation of SVM Using Data Mining Technique to Extract Syllogistic Rules: Exploring the Notion of Explainable AI in Diagnosing CAD. In Proceedings of the International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, 25–28 August 2020.
79. Valentin, S.; Harkotte, M.; Popov, T. Interpreting neural decoding models using grouped model reliance. *PLoS Comput. Biol.* **2020**, *16*, e1007148. [[CrossRef](#)] [[PubMed](#)]
80. Jayaswal, V. Performance Metrics: Confusion Matrix, Precision, Recall, and F1 Score. Medium, towards Data Science. Available online: <https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262> (accessed on 15 February 2024).
81. Silwal, D. Confusion Matrix, Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures. Available online: <https://www.linkedin.com/pulse/confusion-matrix-accuracy-precision-recall-f1-score-measures-silwal> (accessed on 15 February 2024).
82. Bonnet, A. Accuracy vs. Precision vs. Recall in Machine Learning: What is the Difference? Available online: <https://encord.com/blog/classification-metrics-accuracy-precision-recall/> (accessed on 15 February 2024).
83. Wu, J.Y. *Housing Price Prediction Using Support Vector Regression*; San José State University: San José, CA, USA, 2017.
84. Yang, H.; Kang, D.; Hwang, K.; Yang, Z.; Jiang, Y. House Price Prediction with Creative Feature Engineering and Advanced Regression Techniques, NYC Data Science Academy. Available online: <https://nycdatascience.com/blog/student-works/house-price-prediction-with-creative-feature-engineering-and-advanced-regression-techniques/> (accessed on 10 February 2024).
85. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* **2019**, *267*, 1–38. [[CrossRef](#)]
86. Gao, L.; Guan, L. *Interpretability of Machine Learning: Recent Advances and Future Prospects*; IEEE MultiMedia: Washington, DC, USA, 2023; pp. 1–12.
87. Alangari, N.; El Bachir Menai, M.; Mathkour, H.; Almosallam, I. Exploring Evaluation Methods for Interpretable Machine Learning: A Survey. *Information* **2023**, *14*, 469. [[CrossRef](#)]
88. Yekkehkhany, B.; Safari, A.; Homayouni, S.; Hasanlou, M. A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry SAR data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 281–285. [[CrossRef](#)]
89. Nalepa, J.; Kawulok, M. Selecting training sets for Support Vector Machines: A review. *Artif. Intell. Rev.* **2019**, *52*, 857–900. [[CrossRef](#)]
90. Zillow. Why Do Pending Home Sales Fall Through? Available online: <https://www.zillow.com/learn/why-pending-home-sales-fall-through/> (accessed on 15 February 2024).
91. Keys, B.J.; Mulder, P. *Neglected No More: Housing Markets, Mortgage Lending, and Sea Level Rise (No. w27930)*; National Bureau of Economic Research: Cambridge, MA, USA, 2020.
92. Sirmans, G.S.; Benjamin, J. Determinants of market rent. *J. Real Estate Res.* **1991**, *6*, 357–379. [[CrossRef](#)]
93. Yuan, J.; Zheng, X.; You, J.; Skibniewski, M.J. Identifying Critical Factors Influencing the Rents of Public Rental Housing Delivery by PPPs: The Case of Nanjing. *Sustainability* **2017**, *9*, 345. [[CrossRef](#)]
94. Understanding Rental Rates: A Deep Dive into the Influencing Factors. Available online: <https://www.assureshift.in/blog/factors-affecting-rental-prices> (accessed on 20 February 2024).
95. Grybauskas, A.; Pilinkienė, V.; Stundžienė, A. Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. *J. Big Data* **2021**, *8*, 105. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.