Article

Decoding Real Estate Dynamics: A Machine Learning Approach to Housing Price Forecasting

Anas Alsuhaibani [1,*], Mohammed Alswailm [1], Khaled Alnahdi [1] and Abdullah Alghamdi [1]

[1] Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University,11942, Al-Kharj, Saudi Arabia; ah.alsuhaibani@psau.edu.sa, 41050338@std.psau.edu.sa, 441051053@std.psau.edu.sa, 439050509@std.psau.edu.sa

* Correspondence: ah.alsuhaibani@psau.edu.sa

Abstract: This research delves into the pivotal role of Machine Learning (ML) in reshaping the landscape of housing analysis and price forecasting. Recognizing housing as a fundamental aspect of human existence with profound societal and individual implications, the study leverages ML advancements to overcome challenges in comprehensive data analysis and accurate price predictions. The project's focal point is the development of a Random Forest (RF) model that utilizes housing structural details to predict prices, offering valuable insights for investors and real estate stakeholders. Key inquiries include assessing the model's accuracy, identifying influential features, and exploring challenges and limitations in employing an RF model for housing price estimation. The research entails a thorough review of prior studies, a detailed methodology encompassing data collection and model development, and concludes with the presentation of results and a comprehensive dataset report. This investigation not only contributes to advancing housing market understanding but also provides a framework for informed decision-making in the dynamic realm of real estate.

Keywords: keyword 1; keyword 2; keyword 3 (List three to ten pertinent keywords specific to the article; yet reasonably common within the subject discipline.)

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Information* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

1. Introduction

Housing is an essential aspect of human existence. It is of utmost importance in shaping societies and the well-being of individuals and families. Far from being an indispensable and basic human need, housing plays a pivotal role in influencing the stability of countries economically, the dynamics of countries socially, and the quality of life in its general form. With the advent of technological advancements, especially in the field of Machine Learning (ML), unique opportunities to enhance various aspects of the housing sector have increased.

The role of ML has been increasingly integrated into diverse fields, revolutionizing traditional methods and producing unprecedented insights. In the context of housing, ML presents itself as a powerful tool to revolutionize the way we analyze, forecast, and improve various aspects of the housing market. By harnessing the power of data and the right algorithm, ML can enhance the overall housing experience.

In the context of the housing market and its developments, one of the most important markets in this sector, which is growing significantly and has a clear vision, is the Saudi market. The position of the housing sector and real estate activities among Saudi economic activities as a whole is considered influential. This sector is considered the fifth most influential sector in contribution to the Gross Domestic Product (GDP) at a rate of 6.8% [1]. Over time, this sector will become more influential, as the Saudi government aims to raise the sector's contribution to the GDP by 10% by 2030. One of the most prominent initiatives in raising the contribution is the establishment of Sakani program led by the Ministry of Housing which aims to provide housing solutions and products that meet the needs of citizens.

In light of this, housing research holds great promise due to the significant challenges of the housing market and real estate forecasting prices. Although a lot of information is available about the real estate situation and its trends, the issue of obtaining comprehensive data on housing characteristics, analyzing them, and accurately forecasting their prices remains a challenge.

Accordingly, our project aims to build a predictive model for housing prices based on housing structural details. We will use ML techniques to develop a Random Forest (RF) model that can predict housing prices based on different variables related to housing characteristics.

This will contribute to providing a comprehensive overview for investors and those interested in the real estate market about the factors determining housing prices, and we aim to answer the following questions:

- What is the accuracy of using the RF model house prices estimation?
- What features most influence the price estimate?
- What are the challenges and limitations of using a RF model to estimate housing prices?

In the following sections, we will review previous research related to housing price forecasting, examine the approaches used, and the results achieved to shed light on this research. After that, the research methodology will be detailed, including collecting and cleaning the data, choosing the model, training it, and then evaluating it. Finally, the results of this research will be reviewed, accompanied by a report on the collected dataset.

2. Literature Review

2.1. Housing Market

Housing is an indispensable thing in life, just as there are physiological needs such as breathing, drinking water, and eating, housing is an integral and essential part of human life [2].

Because it is an essential part, it must have a market, and because it has a market, there must be some indicators used to measure changes in housing prices and the House Price Index (HPI) appears. This index is used in many countries, such as the United States (US) and the United Kingdom (UK). However, this index is not sufficient to estimate the price of a particular house in the current century because it is an approximate index and provides a summary figure based on previous transactions of buying and selling and not based on specific details about the individual characteristics of the house [3,4]. The HPI obtains previous transactions upon which it relies by reviewing recurring mortgage loan transactions on single-family properties whose properties were purchased and then measures changes in prices over time [4].

In general, certain factors and characteristics affect the value and price of housing, and among these factors are the quality of the building and its spatial characteristics such as location and neighborhood [5]. Location is important and influential in the process of moving residents from one home to another, as municipal services, hospitals, and even shopping centers and workplaces move from one area to another. The quality of the neighborhood is also one of the most important factors that affect the price of a house, and its price may increase despite potential limitations in amenities or space because it is located in a safe neighborhood and satisfactory services are available around it [5]. Furthermore, the environmental conditions of the area, such as the level of pollution and noise, are among the influential factors that may contribute to the increase or decrease in the price of the house [6]. The price is also affected by an increase or decrease due to the characteristics of the house, including the space, the number of rooms, the age of the building, and even the number of floors and the existence of elevators [4].

Aside from the factors affecting house prices, the housing market is a benchmark and an essential element of any viable economy in many countries and has either a positive or a negative impact on the economic growth of a country [2,7]. The construction industry may also prove to be healthy if there is a high level of housing supply [2,7]. This can attract significant investments from individuals who have the required capital, as real estate is viewed as an investment vehicle rather than just a commodity, providing investors with growth opportunities [2,7]. The market is more stable compared to other financial markets and is considered an opportunity for many young people who are starting their careers, as owning a piece of real estate is a symbol that raises one's social status [7]. Finally, housing is not only an economic value but also an asset of social importance [2].

2.2. Machine Learning and Prediction Algorithms

In the age of technology, when talking about the context of artificial intelligence and predictive analysis, the term ML is often mentioned [8]. ML is the science of making computers learn from humans and behave like them, by providing computers with the data necessary to learn [9]. The optimization for the learning process of computers grows independently over time [9]. For the machine to learn and understand the data provided, they must use programmed algorithms to produce more accurate results [10].

ML algorithms are many and although there are some variations of how to group ML algorithms, they can be divided into three broad categories according to their purposes and the way the underlying machine is being taught. These categories are supervised, unsupervised, and semi-supervised learning [10].

For the first category, supervised machine learning algorithms are suitable for two types of problems, namely classification and regression problems. The way to deal with these problems is by dividing the data into two different parts, the first of which is labeled training data, and the second is test data. First, train the selected algorithm with labeled training data. Subsequently, evaluate its performance on the test dataset to determine if it effectively categorizes data into the desired groups or makes accurate predictions for the target [10]. There are three common algorithms for supervised learning. The first algorithm is Simple Linear Regression (SLR). In this type, it is assumed that there is a linear relationship between the dependent variable (the variable we want to predict - Y) and the independent variables (variables that affect the dependent variable - X) [11]. By fitting a regression line between the dependent and non-dependent variables, this linear relationship is created [11]. The second common algorithm, the RF algorithm, is a powerful ensemble learning algorithm that relies on combining multiple decision trees, and since there are many trees, this allows for improving prediction accuracy and avoiding overfitting compared to a single decision tree, because all decision trees are participating in voting or issuing a decision [12]. The decision of whoever receives the largest number of votes from the trees is the final decision [12]. Third, there is the XGBoost algorithm which is a gradient-boosting algorithm, which means it works by gradually building weak models and combining them to create a strong model [12]. The weak models here are decision trees, and these models are trained sequentially to correct previous errors until a robust predictive model is created [12].

As for unsupervised learning, it is used for clustering and feature reduction tasks and is the exact opposite of supervised learning [13]. In this type, there is no trained labeled data, as the algorithm learns the input features on its own and it discovers and analyzes the structure of the data and extracts useful information from it [13]. Clustering is a popular unsupervised learning technique that detects similar groups in data [14]. It aims to group data points based on their similarities or differences and not directly predict the output of unseen inputs [14]. Feature reduction is also a learning technique in unsupervised learning, as this technique allows for to reduction of noise and redundancy in the dataset and finding an approximate version of the dataset using less data [15].

The semi-supervised learning, falls between supervised learning and unsupervised learning, meaning that it is a combination of the two [13,16] Semi-supervised learning is concerned with using labeled and unlabeled data, which means that the algorithm is trained by exploiting the information available in the labeled data and then the trained algorithm is used to generalize classifications to the unlabeled data [16].

Typically, the goal is to take advantage of additional information to improve performance [16]. For example, in clustering methods, there may be knowledge that some data points belong to the same category and this knowledge provides utility for conducting this type of learning [16]. In classification problems, additional data points whose classification is unknown can be used to help process the classification [16]. Finally, ML types are one of the layers that simplify decision-making to choose the appropriate algorithm. They help narrow the scope to reach the characteristics of the data and are the second layer that helps guide the final selection of the appropriate algorithm [17].

2.3. Related Works

In the context of prediction models and housing, few studies have been conducted. Adetunji et al. [18] conducted a study on a housing dataset from Boston city in the US. They retrieved their data from Kaggle platform where the researchers implemented a RF model on the dataset. They found that the model (RF) achieved prediction accuracy of around 90% when comparing predicted house prices to actual prices. This result confirms the effectiveness of this model (RF) in accurately predicting house prices based on the features available in the Boston housing dataset. However, it is important to note that their study did not include features related to housing characteristics such as space, house age, house price, and other amenities, which are more influential factors than economic and social variables.

In a similar vein, Sanyal et al. [10] delved into a Boston housing dataset, using a variety of traditional regression models, including SLR, Polynomial Regression, Ridge Regression, and Lasso Regression. Their findings revealed that Lasso Regression emerged as the best performing model among the regression techniques reaching an accuracy of up to 88%. However, in addition to the fact that the dataset does not include structural features, it is also important to realize that the regression models used in their study, although effective, do not reach the level of advance shown by algorithms such as RF and XGBoost. This is because these advanced algorithms have the capabilities to handle many complex features and interactions, which makes them achieve better predictions than other traditional algorithms.

In the context of advanced algorithms, Abdul-Rahman et al. [19] took a different approach, implementing two advanced algorithms, LightGBM and XGBoost on a housing dataset specific to Kuala Lumpur, Malaysia. This dataset includes some features related to structural house properties, sourced from Kaggle and Google Maps. These advanced algorithms were then compared to traditional regression models, ultimately showing that the XGBoost-based house price prediction model provided the highest level of performance with an accuracy of nearly 91%, outperforming other ML models. However, the structural features in the dataset were not abundant and were limited. It lacked some features, such as the space of the house, the name of the neighborhood, the number of lounges, the width of the street, and some amenities such as elevators, stairs, etc. This limitation in features does not give a clear picture of the impact of amenities on the price of the house, unlike if they were abundant.

Monika et al. [20] conducted a study on forecasting housing prices in different cities in the US and found that the LightGBM model achieved a accuracy of up to 90% according to the metrics used compared to several models such as XGBoost and Support Vector Machine (SVM). However, against the high accuracy, there was a limitation in the number of house records for its dataset as it is only 1460 records and this limitation in number can make it an imprecise measure of prediction.

Chowhaan et al. [21] built several advanced models such as RF, SVM, and XGBoost to predict house prices collected from real estate websites, this dataset contains features such as the number of rooms and location. The researchers found that the RF model was the best-performing model, with an accuracy of 90%, followed by XGBoost, achieving an accuracy of approximately 88%. However, the researchers admit that their study lacks several features and records in the dataset as it is limited compared to other studies.

Despite the progress achieved, a common thread among these studies appears to be the lack of comprehensive datasets, making uncertainty prevalent in knowing the factors that influence the price. Datasets that are diverse in features and rich in records are challenging because of the value they add. Added value provides comprehensive insights into the accuracy and performance of predictive models and into the factors that influence house prices more broadly. In light of these challenges, this research aims to address the existing gap in the literature by providing a solution to the scarcity of comprehensive datasets in the realm of housing price prediction. Recognizing the pivotal role that robust datasets play in refining predictive models, our study endeavors to contribute a meticulously curated dataset that encompasses a diverse array of features and a rich repository of records. By doing so, we seek to not only enhance the precision of our predictive models but also provide a valuable resource for future investigations into the multifaceted factors influencing house prices.

3. Methodology

This section outlines our approach to handling the dataset. It encompasses data collection, cleaning, model training, model evaluation, and concludes with data reporting as shown in Figure 1.

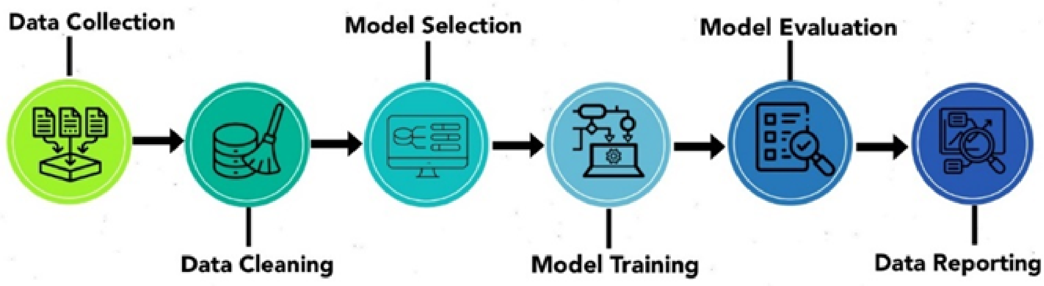

Figure 1. Data Handling Steps

3.1. Data Collection

The dataset utilized in this research was acquired from the Kaggle[1] platform. Kaggle, an online community and resource known for its engagement of data science and ML enthusiasts, provided a unique and comprehensive dataset for this study [22]. The dataset in question encompasses a substantial volume of data, comprising 46,826 individual records denoting houses, alongside 26 distinct columns, each representing specific attributes or characteristics as shown in Table 1.

Notably, it is important to acknowledge that the source of the data, as indicated by the dataset publisher, originates from web scraping activities on the Aqar.sa[2] website. This website is Saudi Arabia's top real estate platform, offering a comprehensive range of property listings in various cities and regions with diverse price ranges and sizes.

[1] https://www.kaggle.com/datasets/reemamuhammed/riyadh-villas-aqar

[2] https://sa.aqar.fm/

Table 1. List of Features in the Dataset

Name	Description
Column1	The index for houses
Front	Describes the direction of the house facade
Rooms	Number of rooms in a house
Lounges	Number of lounges in a house
Bathrooms	Number of bathrooms in a house
StreetWidth	The distance between houses on either side of the road
Stairs	Description of the presence/absence of stairs in the house
PropertyAge	Number of years since it was built
DriverRoom	The presence/absence of a driver room in the house
Tent	The presence/absence of a tent in the house
Patio	The presence/absence of a patio in the house
Kitchen	The presence/absence of a kitchen in the house
OutDoorRoom	The presence/absence of an outdoor room in the house
Garage	The presence/absence of a garage in the house
Duplex	Indicates whether it's a duplex house or not
Space	House area in square meters
Apartments	Number of apartments in a house
MaidRoom	The presence/absence of a maid room in the house
Elevator	The presence/absence of an elevator in the house
Furnished	Indicates whether the house is furnished with all furniture or not
Pool	The presence/absence of a pool in the house
Basement	The presence/absence of a basement in the house
Neighborhood	Name of the neighborhood
Location	Name of the location
SquarePrice	Price per square meter of the house
Price (Target)	Price of the house sold in the market

3.2. Data Cleaning

Prior to selecting and training the model, thorough processing and cleaning of the data were conducted to ensure the accuracy and reliability of the results. The following activities were implemented to clean and prepare the data for this study:

- **Reformatting "neighborhood" Values:** Unified similar names by removing extra spaces using the TRIM feature in Excel. For example, reformatting " النرجس" to "النرجس".
- **Removing Invalid Rows:** Removed rows with neighborhood names that do not exist in reality. Additionally, rows containing street names in the neighborhood field were removed.
- **Handling Missing Values:** Rows containing missing feature values were identified, counted, and then removed from the dataset.
- **Converting Categorical Variables:** Converted categorical variables like "lounges," "bathrooms," and "apartments" to numerical variables for mathematical analysis.
- **Transforming Binary Relationships:** Transformed columns representing binary relationships (e.g., "stairs," "garage," "pool," and "duplex") into binary values (1 or 0) to clearly capture these relationships in the data.
- **Setting Limits for Values:** Set minimum and maximum limits for some columns (e.g., "space," "price," "streetWidth," and "apartments") to address outliers that do not represent reality based on Saudi regulations.
- **Removing Irrelevant Columns:** Removed columns not relevant to the prediction target (price), namely "Column1" and "front." The "squarePrice" column was also removed as it is not considered a target for prediction.

These data cleaning steps were crucial in creating a high-quality dataset for this research, resulting in a dataset comprising 44,109 rows and 23 columns.

3.3. Model Selection

The model selection step is one of the important tasks for making accurate predictions [10]. There are various models available under regression analysis but in this research, the choice fell on the RF Regressor model. This model is a powerful ensemble learning algorithm that combines multiple decision trees to make predictions, making it well-suited for regression tasks, such as houses prices [23]. The algorithm works by creating an ensemble of decision trees, where each tree is trained on a random subset of data and features. The final prediction is determined by averaging the predictions of all individual trees [23]. One of the main advantages of RF Regressor is its ability to handle large and complex datasets [24]. It can effectively handle high-dimensional data with many features [24], which is especially useful for predicting house prices because it often involves considering multiple variables. Furthermore, RF Regressor can capture nonlinear relationships and interactions between variables, allowing it to capture complex patterns in the data [25]. This flexibility allows the model to make accurate predictions, even when relationships between predictors and the target variable are nonlinear or involve complex relationships [25]. Given these advantages, the RF Regressor was considered the most suitable model for this research, and it was trained and evaluated using appropriate performance metrics to evaluate its predictive capabilities in estimating house prices.

3.4. Model Training

Before training the model, some columns must be processed and data must be converted from categorical data to numerical format after it has been analyzed. These are the “location” and “neighborhood” columns because machine learning algorithms are designed to work on numerical data. To aid conversion, a Label Encoder was used. This tool assigns a unique numeric label to each category in a categorical variable. For example, in the "location" column, if the category variables are "شرق الرياض", "غرب الرياض", "جنوب الرياض", "شمال الرياض", and "وسط الرياض", the areas in Riyadh, the Label Encoder will assign labels 0, 1, 2, 3, and 4 for these variables, respectively. After processing has been completed, to train the RF model to predict prices, the dataset is divided into two sets: the training set and the test set. The training set represents 80% of the dataset, while the test set makes up the remaining 20%, and this is common practice [26]. During the training phase, the model learns and feeds from the training sample by analyzing features (such as location, neighborhood, space, etc.) and their corresponding target variable (price). After the training phase is completed, the RF model performance is evaluated using the test set. The test set contains new, unseen data with features but without the target variable (price). The model (RF) generates price predictions based on what it learned during training. These predicted prices are then compared to actual prices in the test set and evaluated using various metrics to test the RF model performance.

3.5. Model Evaluation

The final step is to test and evaluate the RF model predictions for houses prices. Regardless of the model used, no prediction method produces a 100% accurate value [2]. If the future is known absolutely, this is not a prediction, therefore every prediction has a certain percentage of error, and one of the most important criteria that compares the predictive successes of models is the criterion of prediction accuracy [2]. In order to measure the accuracy of the predictive success of the model, the Mean Absolute Percentage Error (MAPE) and R-squared (R^2) criteria were applied in this research. These two criteria estimate the accuracy of the model by analyzing the predicted errors and are calculated using the following evaluation formulas:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100$$

y_t denotes the actual value, $\hat{y}_t$ denotes the predicted value, and n is the number of rows.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

In this formula, y_i represents the actual value, $\hat{y}_t$ represents the predicted value, and $\bar{y}$ is the mean of the actual values.

4. Results

4.1. Performance Evolution

After the two criteria MAPE and (R^2) were used to evaluate the predictive RF model for house prices. The results of the model performance evaluation were revealed as shown in Table 2 and also the prices predicted by the model were compared against the actual prices as shown in Figure 2.

Table 2. Results of Evaluation

Evaluation Metric	Result
MAPE	0.06
R^2	0.948

The results of this model showed its effectiveness in estimating housing prices, as shown in Table 2. The closer the MAPE value is to zero, the more it shows that the predictive model agrees well with the actual values [27]. On the contrary, as the value of R^2 increases and approaches 1, it indicates that the model was a good fit with the actual values [28].

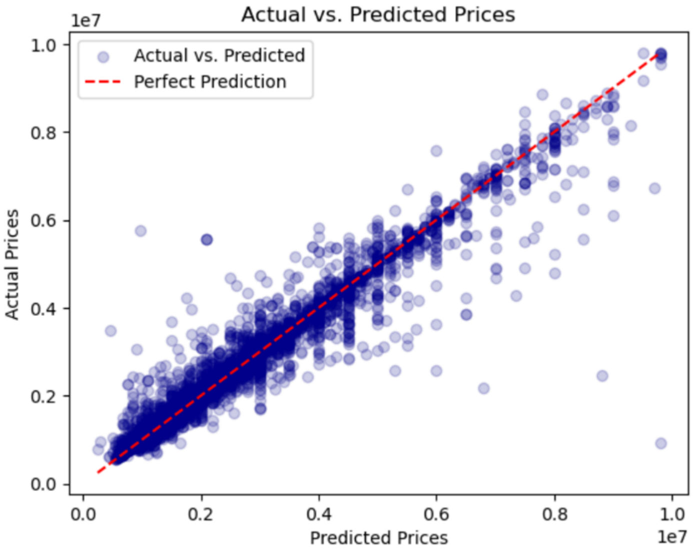

Figure 2. Comparison Between Actual and Predicted Prices

318

4.2. Data Reporting 319

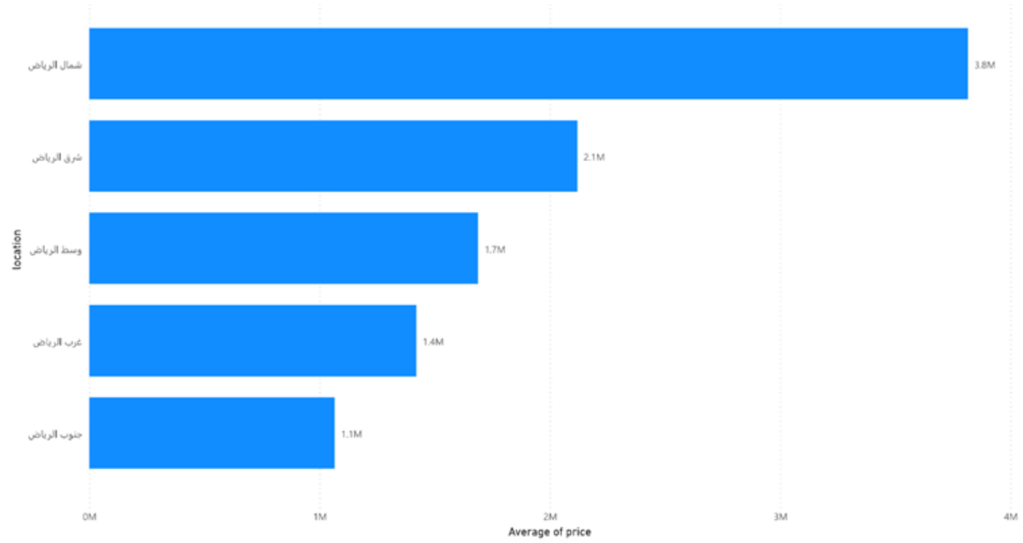

Figure 3. Price Average by Location

It is noted in Figure 3 that the average price in northern Riyadh is higher than its closest competitor (east of Riyadh) by approximately 45%. While the average percentage difference between all other locations (except the north) reaches 19.3%. This relative disparity in price rates shows the extent of the difference in social class between one location and another, as well as in terms of infrastructure and available services. 320–324

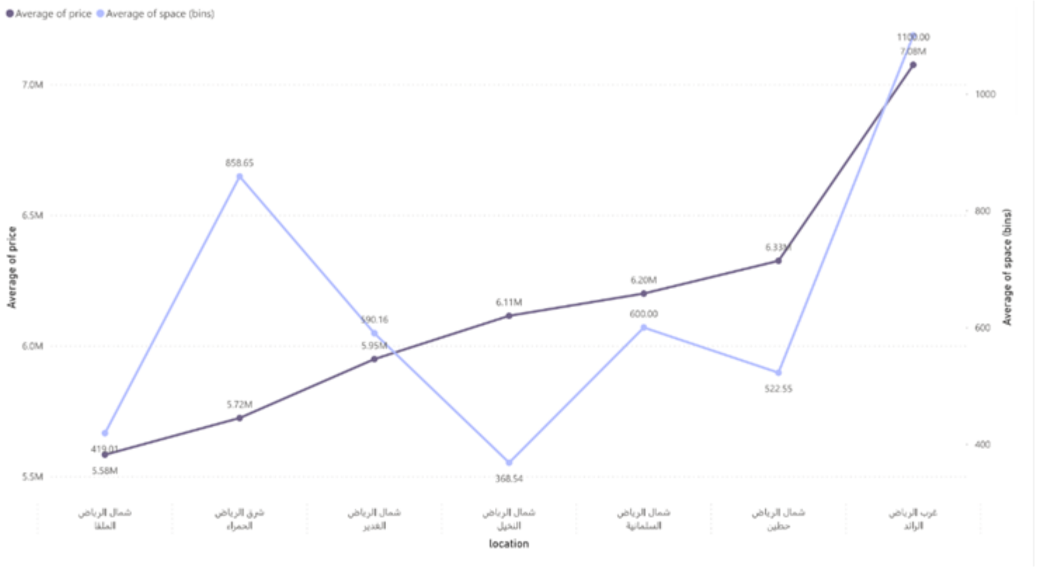

Figure 4. Top Seven Neighborhoods by Average Price

In Figure 4, it appears that 5 (71%) of the top 7 neighborhoods in Riyadh are located in the north, with an average space of about 500 square meters. It is noteworthy that the average space in the remaining percentage of neighborhoods that are located in two different locations (west and east) exceeds the average space in the north by up to 96%. In general terms, this disparity in space may explain the extent of the difference in cultures between the inhabitants of one location and another. 325–330

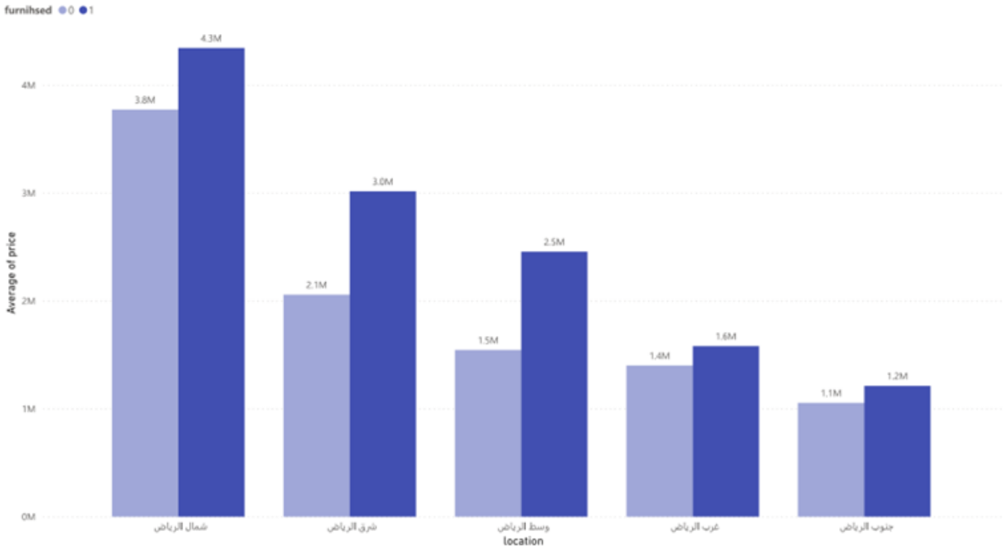

Figure 5. Comparing Average Price of Furnished/Unfurnished by Location

It can be concluded from Figure 5 that the average difference between the prices of furnished and unfurnished houses in (south, west and north of Riyadh) is 72% less than the difference rates in (central and east of Riyadh). This percentage may indicate the interest of investors in the two locations, whose social class may be lower than that of the population of the north and higher than that of the west and south.

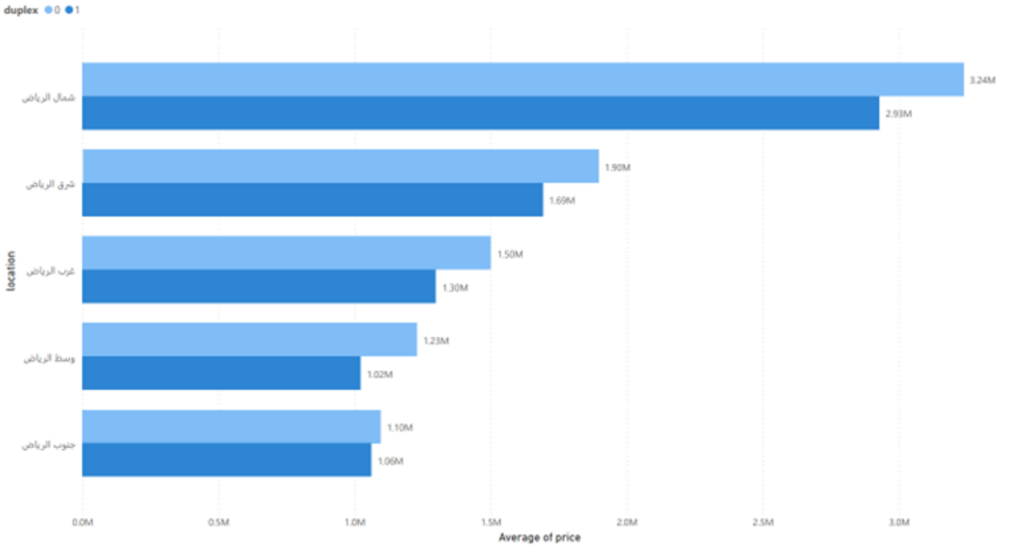

Figure 6. Comparing Average Price of Duplex/Non-duplex by Location

What was found in Figure 6 is that in southern Riyadh, the average price of a duplex house is about 40 thousand riyals less than a house a non-duplex. This difference is considered about 6 times less than the average difference between prices in other locations (235 thousand riyals). The reasons behind this may be that the facilities in non-duplex houses may be limited compared to duplex houses. Furthermore, the issue of low supply and demand in the eastern locations could be one of the most important factors that contributed in making the rate low compared to the rest of the locations.

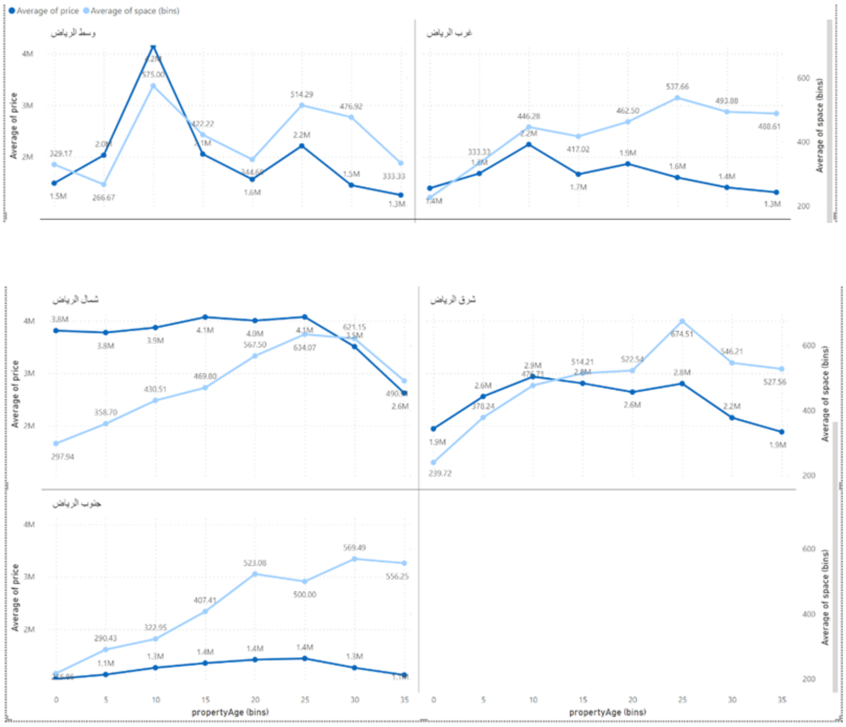

Figure 7. Average Price Space for Each Location by Property Age

What was found from Figure 7 is that the average space of houses that are not more than five years old is approximately 293 square meters. While the average space of houses between 30 and 35 years old is 74% higher than modern houses. This percentage shows that space has a direct relationship with the age of the house.

5. Conclusion

In conclusion, our research seeks to leverage the power of ML, specifically by implementing a RF model, to address complex challenges in housing price forecasting. Recognizing the pivotal role of the housing sector in countries' economies and shaping their societies, our research aims to contribute valuable insights to investors and those interested in this sector.

The use of ML techniques provides a promising way to predict and analyze housing prices based on structural details. Through this research, we sought to answer many questions regarding the accuracy of the model in estimating housing prices, the features affecting its estimation, and finally the challenges and limitations associated with using such models in the context of forecasting housing prices.

References

1. Sakani Report. Available online: https://sakani.sa/sakani-report
2. Soy Temür, A., Akgün, M., Temür, G.: Predicting housing sales in Turkey using ARIMA, LSTM, and hybrid models. Journal of Business Economics and Management. 20, 920–938 (2019). https://doi.org/10.3846/jbem.2019.10190
3. Lu, S., Li, Z., Qin, Z., Yang, X., Goh, R.S.M.: A hybrid regression technique for house prices prediction. In: IEEE International Conference on Industrial Engineering and Engineering Management. pp. 319–323. IEEE Computer Society (2017)
4. Quang, T., Minh, N., Hy, D., Bo, M.: Housing Price Prediction via Improved Machine Learning Techniques. In: Procedia Computer Science. pp. 433–442. Elsevier B.V. (2020)
5. Masri, M.H. b M.@, Nawawi, A.H. b, Sipan, I. b: Review of Building, Locational, Neighbourhood Qualities Affecting House Prices in Malaysia. Procedia Soc Behav Sci. 234, 452–460 (2016). https://doi.org/10.1016/j.sbspro.2016.10.263

6. Cygas, Donatas., Froehner, K.Dieter., Breznikar, A.: Environmental engineering: selected papers. Vol. 3, Sustainable urban development. Roads and Railways. Technologies of Geodesy and Cadastre. Vilnius Gediminas Technical University Press "Technika" (2011)
7. 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 (Switzerland). 11, (2022). https://doi.org/10.3390/electronics11213448
8. Korjo Hwase, T., Fofanah, A.J.: Machine Learning Model Approaches for Price Prediction in Coffee Market using Linear Regression, XGB, and LSTM Techniques. International Journal of Scientific Research in Science and Technology. 8, 10–48. https://doi.org/10.32628/IJSRST
9. Faggella, D.: What is Machine Learning? Available online: https://emerj.com/ai-glossary-terms/what-is-machine-learning/
10. Sanyal, S., Kumar Biswas, S., Das, D., Chakraborty, M., Purkayastha, B.: Boston House Price Prediction Using Regression Models. In: 2022 2nd International Conference on Intelligent Technologies, CONIT 2022. Institute of Electrical and Electronics Engineers Inc. (2022)
11. Begum, Amena Kheya, Nishad Zahid, Zahidur. (2022). Housing Price Prediction with Machine Learning. International Journal of Innovative Technology and Exploring Engineering. 11. 42-46. https://www.ijitee.org/portfolio-item/c97410111322/.
12. Begum, A., Kheya, N.J., Rahman, Md.Z.: Housing Price Prediction with Machine Learning. International Journal of Innovative Technology and Exploring Engineering. 11, 42–46 (2022). https://doi.org/10.35940/ijitee.C9741.0111322
13. Mahesh, B.: Machine Learning Algorithms-A Review. International Journal of Science and Research. (2018). https://doi.org/10.21275/ART20203995
14. Sah, S.: Machine Learning: A Review of Learning Types. (2020). https://doi.org/10.20944/preprints202007.0230.v1
15. Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., Saeed, J.: A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. Journal of Applied Science and Technology Trends. 1, 56–70 (2020). https://doi.org/10.38094/jastt1224
16. van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Mach Learn. 109, 373–440 (2020). https://doi.org/10.1007/s10994-019-05855-6
17. Sala R, Zambetti M, Pirola F, Pinto R: How to select a suitable machine learning algorithm: a feature-based, scope-oriented selection framework.
18. Adetunji, A.B., Akande, O.N., Ajala, F.A., Oyewo, O., Akande, Y.F., Oluwadara, G.: House Price Prediction using Random Forest Machine Learning Technique. In: Procedia Computer Science. pp. 806–813. Elsevier B.V. (2021)
19. Abdul-Rahman, S., Mutalib, S., Alam, S., Nor, M., Zulkifley, H., Ibrahim, M.I.: Advanced Machine Learning Algorithms for House Price Prediction: Case Study in Kuala Lumpur.
20. Monika, R., Nithyasree, J., Valarmathi, V., Hemalakshmi, M.G.R., Prakash, N.B.: House Price Forecasting Using Machine Learning Methods. (2021)
21. Chowhaan, M.J., Nitish, D., Akash, G., Sreevidya, N., Shaik, S.: Machine Learning Approach for House Price Prediction. Asian Journal of Research in Computer Science. 16, 54–61 (2023). https://doi.org/10.9734/ajrcos/2023/v16i2339
22. Kaggle: Kaggle, https://www.kaggle.com/
23. Zhang, L.: Housing Price Prediction Using Machine Learning Algorithm. Journal of World Economy. 2, 18–26 (2023). https://doi.org/10.56397/jwe.2023.09.03
24. Oshiro, T.M., Perez, P.S., Baranauskas, J.A.: How many trees in a random forest? In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 154–168 (2012)
25. Auret, L., Aldrich, C.: Interpretation of nonlinear relationships between process variables by use of random forests. Miner Eng. 35, 27–42 (2012). https://doi.org/10.1016/j.mineng.2012.05.008
26. Ahmed: The Motivation for Train-Test Split, https://medium.com/@nahmed3536/the-motivation-for-train-test-split-2b1837f596c3
27. Montaño Moreno, J.J., Palmer Pol, A., Sesé Abad, A., Cajal Blasco, B.: El índice R-MAPE como medida resistente del ajuste en la previsiońn. Psicothema. 25, 500–506 (2013). https://doi.org/10.7334/psicothema2013.23
28. Ozili, P.K.: The acceptable R-square in empirical modelling for social science research. (2023)