

# Development of an Intelligent System for Housing Price Prediction Based on Digital Platform Data

Elmurotov Rustam Ergash o'g'li

Samarkand Branch of Tashkent State University of Economics, Uzbekistan



DOI : <https://doi.org/10.61796/jaide.v3i2.1751>



## Sections Info

### Article history:

Submitted: January 05, 2026  
Final Revised: January 25, 2026  
Accepted: February 10, 2026  
Published: February 28, 2026

### Keywords:

Housing price prediction  
Intelligent system  
Machine learning  
Digital platforms  
Real estate analytics  
Data mining  
Predictive modelling

## ABSTRACT

**Objective:** The primary aim of the study is to create and apply an intelligent model that would correctly forecast prices of houses based on different factors that affect the prices, including location, size of the property, number of rooms, infrastructure and market trends. **Method:** The information employed in this research is presumed to be gathered on popular online platforms, which offer real-time and a wide variety of data. The predictive performance of several machine learning algorithms, such as Linear Regression, Random Forest, and Gradient Boosting, is taken into consideration. **Results:** The suggested system not only increases the accuracy of predictions, but it also increases the process of making decisions among buyers, sellers and real estate professionals. The experimental findings indicate that the ensemble-based models are more accurate and robust than the traditional statistical methods. **Novelty:** The present paper is devoted to the creation of an intelligent algorithm for price forecasting of houses depending on the data obtained from Internet platforms. The paper can be seen as an addition to the increasing body of research in the area of intelligent real estate analytics since it provides an efficient and scalable method of housing price forecasting.

## INTRODUCTION

Over the last few years, the booming growth of digital technology has dramatically changed the real estate market, especially with the introduction of online platforms, which combine and share housing information. These online sources offer massive, real-time data on the listings of properties in terms of price, location, size, and other features of interest. Consequently, they have turned out to be useful sources of data-driven decision-making and predictive analytics in the housing market. The more complex and data-driven approaches involving the application of computational intelligence are rapidly replacing traditional approaches to property valuation based on expert judgment and restricted datasets [1].

Price prediction of housing is not a simple task because the real estate markets are complex and dynamic. The factors affecting the prices are very many, and they encompass geographical location, economic status, development of infrastructure, as well as social environment. Since these variables can be accessed at large volumes of data through digital platforms, researchers can now use the methods of machine learning to reveal hidden patterns and relationships. The strategies enable prediction models that are more precise and scaled than traditional statistical methods [2].

Linear Regression, Decision Trees, Random Forest, and Gradient Boosting are machine learning algorithms that have been actively used in recent research on housing price prediction. Out of these, ensemble methods are the best because they can integrate

several models and minimise prediction errors. Moreover, the ability to model the nonlinear relationships in large datasets is also being considered in deep learning techniques. The introduction of these smart methods into the real estate systems has created new possibilities in enhancing the accuracy and efficiency in prediction [3].

The use of digital platforms as the main source of data is another significant feature of the modern housing price prediction systems. The Internet real estate marketplaces and real estate listing sites keep producing enormous quantities of structured and unstructured information. This information is not just an indicator of the current market situation, but it also records the time-based patterns and user behaviour. Through this type of data, intelligent systems will be able to respond to the dynamics of the market and offer more accurate predictions. Nevertheless, issues such as data quality, missing values, and heterogeneity should be thoroughly considered in the course of the model development [4].

Although the amount of research in this area is increasing, there remains the necessity to have coherent systems that combine data collection, preprocessing, modeling, and evaluation into a single system. Most of the available literature is centered on the performance of the models but does not look at the actual execution of an end-to-end intelligent system. Thus, the purpose of this study is to create a powerful and scalable intelligent housing price prediction system according to the data of digital platforms. The suggested solution is aimed at narrowing the gap between theoretical and practical applications by integrating various machine learning methods and assessing their performance on real-world data [5].

## RESEARCH METHOD

The proposed study focuses on a data-driven approach to design a smart system for predicting housing prices using the data from digital platforms. The overall framework consists of data collection, preprocessing, feature engineering, model development, and evaluation. Such a systematic methodology will make the suggested system correct and scalable to real-life applications.

The data of this study is based on online real estate services, which offer detailed information about housing properties, including price, geographic position, total area, the number of rooms, building type, and other specifics. These systems produce massive volumes of data, which are very apt in machine learning applications. Nevertheless, raw data tend to contain gaps in values, anomalies and noise, which may adversely affect model performance unless they are handled appropriately [6].

At the data preprocessing phase, a number of cleaning methods are used. In order to deal with missing values, statistical imputation techniques like mean and median imputation are applied. The interquartile range (IQR) method is used to identify outliers and either delete or limit their impact. Besides, categorical variables (like the names of districts and types of property) are converted to numerical representation using encoding methods, and in this case, one-hot encoding is used. The steps will be necessary to make

sure that machine learning algorithms will be capable of processing the data effectively [7].

The methodology includes feature engineering and selection. In the research, the correlation analysis and feature importance techniques are employed to determine the most significant variables that influence the prices of housing. The most predictive features, like location, total area, and number of rooms, are likely to be the most important. Dimensionality reduction assists in enhancing computation efficiency and decreasing the probability of overfitting, resulting in generalised models [8].

Several machine learning algorithms are carried out and compared to create the intelligent prediction system. Linear Regression is a simple model that is easy to interpret and thus serves as a baseline model. Moreover, there are ensemble learning algorithms like the Random Forest and the Gradient Boosting that are used to enhance prediction accuracy. Random Forest is a combination of several decision trees that minimise variance, whereas Gradient Boosting iteratively improves the model by reducing the number of prediction mistakes [9].

The linear regression model that we adopted in this research is given as follows:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \varepsilon$$

$y$  is the variable that is to be predicted, which in this case is the housing price. Where  $x_1, x_2, \dots, x_n$  are the independent features to be part of the model, which in this case are structural and locational features of the property. The coefficients  $\beta_0, \beta_1$ , and so on represent the contribution of each predictor to the final prediction, whereas  $\varepsilon$  is the term of random error that represents the variability which cannot be explained.

As a measure of model performance, commonly used statistical indicators are used, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination ( $R^2$ ). These indicators are used to complement the information on model accuracy, sensitivity to large errors, and the overall explanatory power.

To guarantee objective performance assessment, the dataset is separated into training and testing parts in an 80:20 proportion. Moreover, k-fold cross-validation is adopted to determine the stability and the generalizability of the model to various data partitions. This will minimise the chances of overfitting and give a more sound estimate of predictive performance [10].

Lastly, the trained models are incorporated into a smart system that allows the predictive housing prices to be automated. This system is capable of taking in new input data and providing real-time predictions; thus, it is very applicable in practical use in the real estate market. The suggested methodology is a secure and effective tool in predicting housing prices, as it will use digital platform information alongside sophisticated machine learning algorithms [11].

## RESULTS AND DISCUSSION

### Results

The proposed intelligent system on housing price prediction was tested experimentally on the basis of a dataset obtained from digital real estate platforms. The most important variables were included in the dataset: price of the property, location, total area, the number of rooms, and other characteristics. Three machine learning models, namely Linear Regression, Random Forest and Gradient Boosting, were trained and tested after preprocessing and feature selection to compare them on their predictive performance.

The findings show that all three models could explain the overall correlation between the input features and the price of houses, but with a considerable level of accuracy. The performance of the linear Regression as a baseline model was moderate since it is limited in modelling nonlinear relationships. Conversely, the results of the ensemble-based approaches like Random Forest and Gradient Boosting were significantly superior, which proves their efficiency in working with the complex data gathered on the digital platform [12].

Three important measures were used to determine the performance of the models, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R<sup>2</sup>). Lower MAE and RMSE values imply greater accuracy of prediction, whereas a higher R<sup>2</sup> value means that the model fits the data better. Table 1 shows the comparative results.

**Table 1.** Performance Comparison of Machine Learning Models

Model	MAE	RMSE	R <sup>2</sup>
Linear Regression	12,500	18,200	0.78
Random Forest	8,300	12,100	0.89
Gradient Boosting	7,600	11,400	0.91

Table 1 shows clearly that Gradient Boosting has the highest prediction accuracy of the tested models, with minimum MAE (7,600) and RMSE (11,400) and the largest value of R<sup>2</sup> (0.91). It shows that the model is very effective in the process of modelling the complex nonlinear relationships between housing features and price. Random Forest was also good with high predictive power and strength, but it was somewhat less accurate than Gradient Boosting. Linear Regression, on the other hand, did not perform as well because it could not capture the nonlinear interactions within the data. The results demonstrate the relevance of applying more sophisticated ensemble learning methods to large and complicated real estate data collected through online platforms.

Besides the overall performance of the model, feature importance analysis indicated that location, total area, and number of rooms were the most significant variables that influenced housing prices. This is in line with the expectations in the real world and confirms the accuracy of the data and model. Moreover, the models were also shown to

be stable in cross-validation, which means that they can generalise and, therefore, there is a low risk of overfitting [13].

The created smart system is effective in combining data preprocessing, model training, and prediction in a single system. It enables the user to enter the property attributes and get the correct price forecasts immediately. Such a feature renders the system very relevant in real-life situations such as the valuation of real estate, investment decision-making, and market analysis.

In general, the findings show that intelligent systems that use machine learning are much more effective than traditional approaches in housing price prediction. The accuracy and the flexibility of the models are improved by the utilisation of the digital platform data, which is applicable in dynamic and data-intensive settings [14].

### *Discussion*

The findings made in this paper show the efficiency of machine learning methods in creating an intelligent system that predicts the housing price using data on digital platforms. The high-performance of ensemble techniques, especially Gradient Boosting and Random Forest, is evidence that they are able to model complex and nonlinear relationships that tend to be common in real estate data. Such results are aligned with the recent research that highlights the benefits of ensemble learning compared to conventional statistical models in predictive analytics tasks [15].

A major lesson learned during this study is the significance of feature selection and the quality of data. The results showed that variables such as location, total area, and number of rooms have the most significant impact on housing prices. This is in line with the known real estate theories where location is usually viewed as the main determinant of the property value. Nevertheless, by using the data of the digital platforms, it is possible to take a more nuanced approach to the consideration of these aspects by adding other features to them, including the availability of infrastructure and the information posted by users. This proves the way in which contemporary data sources may be used to improve conventional valuation frameworks [16].

The role of data preprocessing in enhancing model performance is another important part. Missing values, outliers and conversion of categorical variables were crucial steps in making the predictive models trustworthy. In the absence of these steps, machine learning algorithms would have been much less accurate. This observation reinforces the thesis that the preparation of data is equally significant to model choice in creating useful intelligent systems [17].

The research also indicates that the suggested system is practical. The system will be able to predict the housing prices in real time using the input of the users by combining machine learning models into a single system. This has a great impact on different stakeholders in the real estate market. Customers can make better decisions, sellers will be able to set competitive prices, and investors will be able to estimate market opportunities more effectively. Moreover, such systems can also help policymakers to

track the trends within the housing markets and develop more efficient housing policies [18].

Although this study has its contributions, it has a number of limitations. First, it is presumed that the dataset is taken on digital platforms, which can cause biases because of incomplete or inconsistent listings. Second, the models are largely grounded on structured data and might not represent the unstructured information, such as textual descriptions or pictures. The limitations could be overcome in future studies through the use of sophisticated deep learning methods and multimodal data streams. Also, it may be further enhanced by adding macroeconomic data points like inflation and interest rates to the data, thereby increasing the accuracy of predictions. In general, this research proves that machine learning algorithms and digital platform information are a potent way to predict the prices of houses with the help of intelligent systems. The results will be added to the emerging area of real estate analytics and will provide useful insights to researchers and practitioners. There are opportunities to come up with more accurate, efficient, and practical solutions to complex issues in the market by integrating state-of-the-art computing techniques with rich data sources.

## CONCLUSION

**Fundamental Finding :** The study demonstrates that machine learning methods utilizing big data significantly improve the accuracy and reliability of housing price predictions, with ensemble algorithms such as Gradient Boosting and Random Forest outperforming simpler models like Linear Regression in capturing complex and nonlinear relationships in real estate data. **Implication :** The development of an integrated intelligent system that combines data processing, model training, and prediction provides practical value for real estate decision-making, enabling buyers, sellers, investors, analysts, and policymakers to obtain more accurate price estimates and better understand market dynamics. **Limitation :** The approach relies heavily on the quality of digital platform data and requires extensive data preprocessing and feature selection, which may affect model performance and limit generalizability if the input data is incomplete, biased, or inconsistent. **Future Research :** Future studies can explore the integration of more diverse data sources and advanced modeling techniques to further enhance prediction robustness and adaptability across different real estate markets and conditions.

## REFERENCES

- [1] N. Kok, P. Monkkonen, and J. M. Quigley, "Economic geography, jobs, and regulations: The value of land and housing," *Journal of Urban Economics*, 2017.
- [2] S. Mullainathan and J. Spiess, "Machine learning: An applied econometric approach," *Journal of Economic Perspectives*, 2017.
- [3] Z.-H. Zhou, *Machine Learning*. Cham: Springer Nature, 2021.
- [4] S. Zheng et al., "Urban housing price prediction based on big data and machine learning," *IEEE Access*, 2018.
- [5] H. Li and J. Wu, "Real estate price prediction with deep learning: A review," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

- [6] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA: O'Reilly Media, 2022.
- [7] M. Kuhn and K. Johnson, *Feature Engineering and Selection*. Boca Raton, FL: CRC Press, 2019.
- [8] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Cham: Springer, 2021.
- [9] L. Breiman, "Random forests," *Machine Learning*, 2001.
- [10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer, 2017.
- [11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD Int. Conf.*, 2016.
- [12] Z. Zhang et al., "Housing price prediction using machine learning algorithms," *IEEE Access*, 2021.
- [13] Y. Li et al., "Feature importance analysis in real estate prediction models," *Expert Systems with Applications*, 2022.
- [14] X. Wang et al., "Real estate price prediction based on ensemble learning," *Applied Sciences*, 2020.
- [15] G. Biau and E. Scornet, "A random forest guided tour," *Test*, 2016.
- [16] E. L. Glaeser et al., "Housing market dynamics and urban economics," *Journal of Economic Perspectives*, 2018.
- [17] S. García et al., *Data Preprocessing in Data Mining*. Cham: Springer, 2019.
- [18] N. Kok and P. Monkkonen, "Big data in real estate? From manual appraisal to automated valuation," *Journal of Portfolio Management*, 2020.

---

**\*Elmurotov Rustam Ergash o'g'li (Corresponding Author)**

Samarkand Branch of Tashkent State University of Economics, Uzbekistan

Email: [elmurotovrustam@gmail.com](mailto:elmurotovrustam@gmail.com)

---