

Midocean university

Financial Technology

Artificial Intelligence for Real Estate Valuation : Improving Property Price Prediction Models in the Saudi Housing Market

Prepared by :

Mahabah Turdi

Supervised by :

Fayez Jarad



Table of Contents

Introduction.....	
Problem Statement.....	
Significance of the study.....	
Research Objective.....	
Research Questions.....	
Research Hypotheses.....	
Literature Review.....	
Research Gap.....	
Theoretical Framework.....	
Research Methodology.....	
Population and sample of the study.....	
Research Instrument.....	
Statistical Analysis Methods.....	
Study Limitations.....	
References.....	

Introduction

The real estate sector is considered one of the most important contributors to economic growth in many countries. It plays a major role in attracting investments, creating job opportunities, and supporting overall economic development. In Saudi Arabia, the real estate market has witnessed significant growth in recent years, mainly due to population increase, rapid urbanization, and government initiatives aimed at increasing homeownership. The housing sector, in particular, has become a key focus area, making accurate property valuation more important than ever.

Property valuation is essential for different stakeholders such as investors, banks, developers, and policymakers. Traditionally, real estate valuation has relied on methods like market comparison and expert judgment. While these approaches are widely used, they are often time-consuming and may not always provide consistent or highly accurate results. This is mainly because they depend on human experience and historical data, which can sometimes lead to errors or bias, especially in fast-changing markets.

With the rapid development of Artificial Intelligence and Machine Learning, new opportunities have emerged to improve how property values are estimated. These technologies make it possible to analyze large amounts of data and identify patterns that are difficult to detect using traditional methods. In real estate, this has led to the development of models that can predict property prices based on different factors such as location, size, number of rooms, and market conditions.

Using artificial intelligence in real estate valuation offers several advantages. It can improve the accuracy of predictions by capturing complex relationships between different variables. It also reduces the time required to perform valuations, making the process more efficient. In addition, it helps in reducing human bias, which leads to more consistent results. Because of these benefits, many financial institutions have started to rely on these models, especially in areas such as mortgage lending and investment analysis.

However, despite the global growth in using machine learning for property valuation, its application in the Saudi housing market still needs more attention.

The Saudi market has its own unique characteristics, including specific regulations, economic conditions, and market behavior, which may affect how these models perform. In addition, there is still limited research that focuses specifically on improving property price prediction models within the Saudi context.

Therefore, this study aims to explore how artificial intelligence can be used to improve real estate valuation, with a focus on predicting property prices in the Saudi housing market. The study will examine different machine learning models and evaluate their performance in order to identify the most accurate approach. This research is expected to provide useful insights for both academic purposes and practical applications in the real estate sector.

Problem Statement

The real estate sector in Saudi Arabia has experienced rapid growth in recent years, driven by population expansion and government initiatives aimed at increasing homeownership. This growth has led to a higher demand for accurate and efficient property valuation, particularly in the housing market. However, in practice, real estate valuation in Saudi Arabia still relies largely on traditional approaches such as comparative market analysis and expert judgment.

These conventional methods, while widely used, often suffer from several limitations, including reliance on subjective human judgment, limited use of large-scale data, and potential inconsistencies in valuation outcomes. As a result, there is a noticeable gap between the current valuation practices and the desired level of accuracy, efficiency, and consistency required in a rapidly evolving real estate market.

At the same time, recent developments in Artificial Intelligence and Machine Learning have demonstrated strong potential in improving prediction accuracy across various sectors, including real estate. Several studies have shown that machine learning models can outperform traditional valuation methods by

utilizing large datasets and identifying complex relationships between property characteristics and prices. Despite this, the adoption and empirical evaluation of such models in the Saudi housing market remain limited.

Accordingly, this situation raises several important research questions:

- To what extent can machine learning models improve the accuracy of real estate valuation in the Saudi housing market?
- Which machine learning techniques are most effective in predicting property prices in this context?

How do these models compare to traditional valuation methods in terms of -accuracy and reliability?

Therefore, the core problem of this study lies in addressing the gap between traditional real estate valuation methods and the potential of data-driven approaches, by evaluating the effectiveness of artificial intelligence techniques in improving property price prediction models within the Saudi housing market.

Significance of the study

1. Academic Significance

This study contributes to the academic literature by examining the application of Artificial Intelligence and Machine Learning in real estate valuation within the Saudi context. While previous studies have widely explored machine learning models for property price prediction in different international markets, there is still limited research focusing specifically on the Saudi housing market.

By addressing this gap, the study provides empirical evidence on the effectiveness of machine learning techniques in improving valuation accuracy in a developing real estate market. In addition, it offers a comparative analysis of different predictive models, which can serve as a useful reference for future academic research in the fields of finance, real estate, and financial technology.

2. Practical Significance

From a practical perspective, this study is important for several stakeholders involved in the real estate sector in Saudi Arabia. As highlighted in the problem statement, traditional valuation methods may not always provide accurate or consistent results. Therefore, improving property price prediction models using artificial intelligence can enhance the quality of decision-making in the market.

The findings of this study can support investors in making better investment decisions, assist financial institutions in improving mortgage lending assessments, and help real estate developers better understand market trends. In addition, adopting more accurate and data-driven valuation models can contribute to increasing transparency and efficiency in the Saudi housing market.

Overall, this study supports the transition toward more advanced and technology-driven practices in real estate valuation, which aligns with the ongoing development and modernization of the sector.

Research Objective

General Objective:

To examine the role of Artificial Intelligence and Machine Learning in improving real estate valuation by enhancing property price prediction models in the Saudi housing market.

Specific Objectives:

- To identify the key factors influencing property prices in the Saudi housing market.**
- To apply different machine learning models to predict real estate prices using available datasets.**
- To compare the performance of selected machine learning models in terms of prediction accuracy and reliability.**

- To evaluate the effectiveness of machine learning techniques in improving real estate valuation compared to traditional methods.**
- To provide recommendations for adopting data-driven valuation models in the Saudi real estate sector.**

Research Questions

**What are the key factors influencing property prices in the Saudi housing
-market?**

**To what extent can Machine Learning models be used to predict real estate
-prices accurately?**

**Which machine learning models provide the highest prediction accuracy in real
-estate valuation?**

**How do machine learning-based valuation models compare to traditional real
methods in terms of accuracy and reliability? -estate valuation**

**How can Artificial Intelligence be effectively applied to improve property price
the Saudi housing market? -prediction models in**

Research Hypotheses

Main Hypothesis:

- Machine Learning models significantly improve the accuracy of real estate valuation compared to traditional methods in the Saudi housing market.
-

Sub-Hypotheses:

- Machine learning models provide more accurate property price predictions than traditional regression models.
- There are significant differences in prediction accuracy among different machine learning models.
- Advanced models such as Random Forest and Gradient Boosting achieve higher accuracy compared to simpler models.
- The use of data-driven models reduces errors in property valuation.

Literature Review

1. Antipov & Pokryshevskaya (2012) – Machine Learning for Housing Prices

Reference (APA):

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random Forest for valuation and a comparison with hedonic pricing. *Expert Systems with Applications*, 39(9), 8040–8048.

Objective:

To compare traditional hedonic pricing models with Random Forest models in predicting residential property prices.

Methodology:

- Dataset: Residential apartments
- Model: Hedonic regression vs Random Forest
- Evaluation: Prediction error comparison

Key Findings:

Random Forest significantly outperformed hedonic regression in prediction accuracy and handled nonlinear relationships better.

Relation to Your Study:

Directly supports your use of AI models for property valuation in Saudi Arabia.

Critical Analysis:

While the study proves ML superiority, it is based on a non-Middle Eastern market, which limits generalizability to Saudi Arabia due to differences in market structure and pricing behavior.

2. Park & Bae (2015) – Real Estate Valuation Using Machine Learning

Reference (APA):

Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction. *Expert Systems with Applications*, 42(6), 2928–2937.

Objective:

To evaluate the performance of machine learning algorithms in predicting housing prices.

Methodology:

-Algorithms: ANN, SVM, Regression models

-Data: Residential housing dataset

-Evaluation: RMSE and MAE

Key Findings:

Machine learning models, especially neural networks, provided higher prediction accuracy than traditional regression models.

Relation to Your Study:

Supports your methodological choice of AI-based valuation models.

Critical Analysis:

The study demonstrates strong predictive performance but lacks interpretability, which is critical in real estate decision-making contexts such as investment evaluation.

3. Bao et al. (2014) – Nonlinear Prediction Models

Reference (APA):

Bao, H., et al. (2014). Prediction of real estate price based on SVM and neural network. *Journal of Systems Science and Complexity*, 27(4), 675–688.

Objective:

To improve real estate price prediction using nonlinear AI models.

Methodology:

-Models: SVM, Neural Networks

-Comparative analysis

-Historical housing data

Key Findings:

Nonlinear models outperform linear regression models in capturing complex housing price dynamics.

Relation to Your Study:

Supports your justification for using advanced AI techniques in Saudi housing market prediction.

Critical Analysis:

The study does not incorporate macroeconomic or policy-driven variables, which are highly relevant in the Saudi housing market (e.g., Vision 2030 initiatives).

4. Saudi Housing Market Context (Ministry of Housing / Vision 2030 Reports)

Reference (APA):

Saudi Ministry of Municipal and Rural Affairs and Housing. (2023). *Housing program annual report*. Riyadh, Saudi Arabia.

Objective:

To analyze housing supply, affordability, and demand trends in Saudi Arabia under Vision 2030.

Methodology:

Descriptive statistical analysis of national housing data and policy evaluation.

Key Findings:

The Saudi housing market is experiencing rapid transformation driven by government initiatives, increasing demand, and urban expansion.

Relation to Your Study:

Provides contextual grounding for applying AI models in a rapidly evolving and policy-driven market.

Critical Analysis:

While highly relevant, the report lacks predictive modeling and does not leverage AI techniques, creating a clear research gap that your study addresses.

Research Gap

Despite the growing body of literature on artificial intelligence applications in real estate valuation, several important gaps remain. First, most existing studies have been conducted in developed or non-regional markets such as the United States, Europe, and East Asia, which limits their applicability to emerging markets like Saudi Arabia. The Saudi housing market has unique characteristics influenced by government policies, Vision 2030 initiatives, and rapid urban development, which are not fully captured in previous models.

Second, although machine learning models such as Random Forest, SVM, and neural networks have demonstrated strong predictive performance, many studies focus primarily on prediction accuracy while neglecting model interpretability, which is essential for investment and policy decision-making in real estate.

Third, limited research integrates macroeconomic and policy-driven variables (such as mortgage regulations, housing programs, and urban development projects) into AI-based valuation models. Most studies rely mainly on property-level attributes, ignoring broader market dynamics.

Finally, there is a lack of studies that specifically focus on the Saudi housing market using advanced AI techniques for property price prediction. This creates a clear opportunity for this study to develop and evaluate AI-based valuation models tailored to the Saudi context.

Therefore, this research addresses these gaps by applying machine learning models to Saudi housing data while incorporating market-specific characteristics to improve prediction accuracy and relevance.

Theoretical Framework

The theoretical framework of this study is based on the integration of traditional real estate valuation principles with modern data-driven approaches, particularly those related to Artificial Intelligence and Machine Learning. Real estate valuation has historically relied on established economic and financial theories, including supply and demand dynamics, hedonic pricing theory, and market comparison approaches. These traditional frameworks assume that property values are influenced by a set of observable characteristics such as location, size, quality, and surrounding amenities.

One of the most widely used theoretical foundations in real estate valuation is the Hedonic Pricing Model, which suggests that the price of a property is determined by its individual attributes. According to this model, each characteristic of a property contributes to its overall value, and the total price reflects the combined effect of these attributes. While this approach provides a structured way to understand price formation, it often assumes linear relationships between variables, which may not fully capture the complexity of real-world markets.

In contrast, machine learning models offer a more flexible and advanced approach by allowing the analysis of nonlinear relationships and interactions among

variables. These models do not rely on strict assumptions about data distribution, making them more suitable for handling large and complex datasets. As a result, they can identify hidden patterns and provide more accurate predictions compared to traditional statistical methods.

This study builds on the assumption that property prices are influenced by multiple interrelated factors and that advanced analytical techniques can better capture these relationships. The framework therefore combines traditional valuation concepts with machine learning techniques to evaluate how different models can improve prediction accuracy. It also considers the role of data quality, feature selection, and model performance evaluation in determining the effectiveness of real estate valuation models.

Furthermore, the study is grounded in the concept of Automated Valuation Models (AVMs), which represent the practical application of artificial intelligence in real estate. AVMs use statistical and machine learning techniques to estimate property values automatically, offering a faster and more consistent alternative to manual appraisal methods. By integrating these theoretical perspectives, the study aims to provide a comprehensive understanding of how artificial intelligence can enhance real estate valuation practices in the Saudi housing market.

Research Methodology

This study adopts a quantitative research approach to examine the role of artificial intelligence in improving real estate valuation. The quantitative method is appropriate for this research as it allows for the analysis of numerical data and the evaluation of relationships between variables using statistical and computational techniques. The study focuses on developing and comparing predictive models to assess their performance in estimating property prices.

The research design is based on a comparative and predictive approach. It involves applying multiple machine learning models to a real estate dataset and comparing their performance in terms of prediction accuracy and reliability. This approach enables the identification of the most effective model for real estate valuation within the context of the Saudi housing market.

The data used in this study will be obtained from a publicly available real estate dataset, which includes variables such as property size, number of rooms, location, and price. These variables represent the key factors that influence property valuation and are commonly used in both traditional and modern valuation models. The dataset will be preprocessed to ensure data quality, including

handling missing values, removing outliers, and normalizing variables where necessary.

Several machine learning models will be applied in this study, including Linear Regression, Random Forest, and Gradient Boosting. These models are selected based on their widespread use in previous studies and their ability to handle different types of data and relationships. The dataset will be divided into training and testing sets to evaluate model performance and avoid overfitting.

To assess the accuracy of the models, the study will use standard evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics provide a quantitative measure of how close the predicted values are to the actual property prices. The results will then be compared across different models to determine which approach offers the best performance.

Finally, the findings of the study will be analyzed and interpreted in relation to the research objectives and questions. The study will also discuss the implications of the results for real estate valuation practices and provide recommendations for future research and practical applications.

Population and samples of the study

The population of this study consists of residential real estate properties in the Saudi housing market, including major cities such as Riyadh, Jeddah, and Dammam. The study focuses on property listings that include key attributes such as location, size, number of bedrooms, property type, age, and transaction price.

The sample will be drawn from available housing transaction datasets and real estate listings collected from official platforms and real estate databases. A stratified sampling approach may be used to ensure representation across different cities, property types, and price segments. The final sample size will depend on data availability and data cleaning procedures but is expected to include several thousand property records to ensure model reliability.

Research Instrument

The primary research instrument in this study is a structured real estate dataset. The dataset includes quantitative and categorical variables such as property size, location, number of rooms, property age, amenities, and historical sale prices. In addition, data preprocessing tools and feature engineering techniques will be used to prepare the dataset for machine learning models. No questionnaires or surveys are required, as the study relies entirely on secondary data sources and computational modeling.

Statistical Analysis Methods

This study will employ machine learning and statistical techniques for real estate price prediction. The main methods include:

-Multiple Linear Regression as a baseline model

**Machine Learning models such as Random Forest, Gradient Boosting, and
-XGBoost**

**Model evaluation using metrics such as Mean Absolute Error (MAE), Root
(RMSE), and R-squared -Mean Squared Error**

**Data preprocessing techniques including normalization, encoding, and
-handling missing values**

-Cross-validation to ensure model robustness and avoid overfitting

**The performance of different models will be compared to identify the most
accurate approach for predicting housing prices in Saudi Arabia.**

Study Limitations

This study has several limitations. First, the availability and quality of real estate data in the Saudi market may be inconsistent or incomplete. Second, some important variables such as neighborhood quality or developer reputation may not be fully captured in structured datasets. Third, the models may be affected by market volatility and external economic factors that are not included in the dataset. Finally, machine learning models may suffer from overfitting or limited generalization when applied to unseen data or different market conditions.

References

- Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random Forest for valuation and a comparison with hedonic pricing. *Expert Systems with Applications*, 39(9), 8040–8048.
- Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction. *Expert Systems with Applications*, 42(6), 2928–2937.
- Bao, H., et al. (2014). Prediction of real estate price based on support vector machine and neural network. *Journal of Systems Science and Complexity*, 27(4), 675–688.
- Saudi Ministry of Municipal and Rural Affairs and Housing. (2023). *Housing program annual report*. Riyadh, Saudi Arabia.