QUBAHAN ACADEMIC JOURNAL
VOL. 5, NO. 4, November 2025
https://doi.org/10.48161/qaj.v5n4a1900

Forecasting Saudi Real Estate Market Trends: A Time-Series Based Comparative Study of 2023 Saudi Real Estate Transactional Data

Abdullah A. Aldaeej [1*]

[1] Department of Management Information Systems, Imam Abdulrahman Bin Faisal University, Dammam 34221, Saudi Arabia.

* **Corresponding author:** aaaldaeej@iau.edu.sa.

ABSTRACT: This study presents a structured comparative evaluation of Saudi Arabia's real estate market by applying two core time-series forecasting models: ARIMA and a modified SARIMA model with seasonal adjustment. The dataset, sourced from the Saudi Ministry of Justice, comprises over 10,000 real estate transactions collected weekly across five major cities during 2023, including price, size, transaction volume, and category features. The proposed modified SARIMA model incorporates two key enhancements: (1) dynamic parameter tuning using grid search optimized via AIC and BIC, and (2) an extensible hybrid structure that supports integration with machine learning models for residual learning. Forecasting performance was evaluated using standard metrics. ARIMA achieved an RMSE of 599,867.54 and a MAPE of 35.81%, outperforming modified SARIMA which recorded an RMSE of 609,942.58 and a MAPE of 40.64%. Despite its slightly lower accuracy, the modified SARIMA demonstrated a better statistical fit with AIC and BIC scores of 10.00 and 14.76, respectively. These findings support real-world applications in investment planning, infrastructure allocation by city planners, and regulatory oversight to stabilize property markets.

Keywords: predictive modeling, Saudi real estate, market trends, forecasting accuracy, machine learning.

I. INTRODUCTION

Real estate industry is one of the most important sectors in any country. It is one of the most popular fields for investment and a barometer of the state of the economy. The real estate industry in KSA has grown significantly in the last decade as a result of population growth and urbanization and government efforts to diversify the economy. In response, the Saudi government launched the Vision 2030 plan, which highlights the significance of the real estate sector in the sustainable development of the nation's economy. This plan includes very large-scale projects like the NEOM, the Red Sea Project, and Qiddiya [1] that are expected to revolutionize the real estate industry in the country. Therefore, there is a need to comprehend and forecast real estate prices in such a volatile market to facilitate adequate decision making for investors and policymakers.

Nevertheless, Saudi real estate market is gradually developing and has all the possibilities to become one of the leaders in the region; however, it is rather difficult to forecast real estate prices. Volatility in real estate prices can also be attributed to seasonal variations, government activities, demographic factors, and geographical differences. The conventional methods for forecasting prices, for instance the simple linear regression models, are inadequate in capturing the interactions between these variables and the complexity of the relationships present [2, 3]. Also, there are a lot of barriers to access data, and data quality and availability remain a huge problem. Sometimes the data may contain errors or some information may be missing and this may lead to wrong forecasts that may harm the investment decisions of the investors and policies made by the policymakers [4].

Furthermore, it is important to note that Saudi Arabian market differs from other countries with peculiar cultural, economic and legal factors that affect performance of real estate venture. This uniqueness means it requires the building of specific models of prediction that reflect the Saudi market environment to the dot. There are several obstacles that may be considered in the development of these approaches, and it is now apparent that the advancement of sophisticated machine learning algorithms could present a unique solution to these issues [5]. Nonetheless, their practical implementation with the help of these algorithms involves the consideration of several factors such as the pre-processing of data, feature extraction, selection of an appropriate model, and validation. To tackle these challenges, this study will use a rich dataset containing revenue records of Saudi real estate for 2023 and employ state-of-art predictive analytics. In doing so, we aim at facilitating the detection of underlying market patterns and trends in the Saudi context, enhancing the reliability of the forecast on prices in the real estate market and generating useful insights for involved stakeholders [6].

The rationale for this research is due to the fact that the real estate sector plays a very crucial role in terms of contribution to the KSA's economy as well as due to the paramount importance to forecast the accurate prices in this regard [7]. In light of this, developing a reliable tool and / or methodology for estimating real estate price has become indispensably pertinent given that the country is steadily progressing in the attainment of its Vision 2030 plan. It helps investors to make better decisions that will help them make positive returns, scientist to advise policy makers on what the possible outcomes of certain decisions might be, and it helps the stakeholders to understand the movement of their stocks. Furthermore, recent development of specific branch of artificial intelligence, machine learning, and data analytics, poses a chance to reimplement concept of traditional predictive models within a smarter way, and with better precision and more effectual recommendations. This research is focused on benefiting from the concepts of previous traditional based practices and providing a detailed, efficient mechanism for real estate market price prediction in Saudi Arabia, which can reduce the gap between conventional methodology and new methodologies [8, 9].

The primary research question relevant to this research is the following: what are the barriers towards correctly modeling the real estate prices in Saudi Arabia using transaction data from the year 2023? These traditional models cannot capture all the compound entities in the market forward due to their inability to incorporate interactions and non-linear dependencies among factors [10]. Moreover, the Saudi market environment of real estate has attributes which are different in some degree from those of other countries, including continued growth of urbanization, governmental actions, and cultural factors that also increase uncertainty. These challenges are further compounded by the absence of sophisticated, comprehensive and accurate data resulting in undeveloped forecast information that complicates the decision-making process [11]. The main objective of this study will be to enhance the understanding of the real estate pricing determinants and then create and test more accurate and efficient predictive models based on the machine learning methods. Thus, the study aims at overcoming these challenges in order to create a useful tool for investors/ investors, policy makers, and other stakeholders operating in the Saudi Arabia real estate market.

Unlike conventional SARIMA implementations, our modified SARIMA model introduces a dual enhancement: (i) a hybrid structure that combines seasonal-linear modeling with machine learning-based residual correction, and (ii) a data-driven parameter tuning approach via grid search optimized on AIC/BIC criteria. This integration not only improves predictive accuracy but also adapts the model to the Saudi real estate market's unique dynamics, offering a novel contribution to existing forecasting methodologies.

1. *RESEARCH OBJECTIVE AND NOVELTY*

This study addresses key gaps in the forecasting of real estate prices by proposing an enhanced SARIMA-based framework customized for the Saudi real estate market. The primary objective of this research is to evaluate and improve the forecasting accuracy of SARIMA models for post-pandemic real estate pricing data in Saudi Arabia, while also assessing the practical implications of predicted trends for urban development, investment planning, and policy interventions.

2. RESEARCH QUESTIONS

- RQ1: What data preprocessing challenges affect the accuracy of real estate price modeling using transactional records from Saudi Arabia?
- RQ2: How do the ARIMA and modified SARIMA models compare in terms of predictive accuracy, model fit, and forecasting reliability on 2023 Saudi real estate data?
- RQ3: Can the SARIMA model be effectively enhanced using a hybrid structure and dynamic parameter tuning to better capture market complexity and seasonality?
- RQ4: What are the implications of forecasted real estate trends for urban planning, infrastructure development, and investor strategy?

The novelty of this research lies in the development of a Modified SARIMA model that addresses the limitations of traditional time-series forecasting in the context of the Saudi real estate market. This enhanced model introduces two key innovations. First, it employs a dynamic parameter tuning process using grid search guided by Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), ensuring optimal selection of seasonal and non-seasonal parameters. This eliminates the need for manual trial-and-error and enhances the model's adaptability to real-world data. Second, the model is designed with a hybrid architecture in mind, allowing integration with machine learning algorithms to capture complex, non-linear residual patterns that SARIMA alone cannot model effectively. This hybrid structure bridges the gap between statistical and AI-based forecasting, enabling more accurate and robust predictions. Together, these enhancements make the proposed approach uniquely suited to handle the seasonal trends, volatility, and structural complexity of real estate transactions in Saudi Arabia something not adequately addressed in existing literature.

2.1 Problem Formulation 1: Data Quality and Preprocessing

The feature vectors extracted from real estate transactions various data sets may include missing values, anomalies, and variations that can distort the outputs of the predictive models. The data cleaning and normalization process should be implemented properly to enhance the data set with better quality and more suitable formats for data analysis. The rationale for doing this is to provide an optimal step by step approach to implementing how to handle missing data, how to detect or handle outliers as well as maintain data consistency. Let $X \in \mathbb{R}^{n\times m}$ be the dataset matrix, where n is the number of transactions and m is the number of features. Let x_i represent the i-th row of X, corresponding to the i-th transaction.

2.1.1 Handling Missing Data

Define the missing data indicator matrix $\mathbf{M} \in \{0,1\}^{n\times m}$, where $M_{ij} = 1$ if X_{ij} is missing, and $M_{ij} = 0$ otherwise. The target is to fill missing values with an appropriate technique including mean imputation or the k nearest neighbor imputation (KNN impairment):

$$X_{ij} = \begin{cases} X_{ij}, & \text{amp; if } M_{ij} = 0 \\ \dfrac{\sum_{k=1}^{n} X_{kj}\left(1 - M_{kj}\right)}{\sum_{k=1}^{n}\left(1 - M_{kj}\right)}, & \text{amp; if } M_{ij} = 1 \text{ (mean imputation)} \end{cases} \tag{1}$$

2.1.2 Detecting and Correcting Outliers

Use the interquartile range (IQR) method to detect outliers. Define the IQR for each feature j as $IQR_j = Q3_j - Q1_j$, where $Q3_j$ and $Q1_j$ are the third and first quartiles of feature j. An outlier detection criterion can be:

$$\text{Outlier}_{ij} = \begin{cases} 1 \\ 0 \end{cases}, \frac{\text{if Xij}}{\text{otherwise}} < \text{Q1j} - 1.5 \cdot \text{IQRj or Xij} > \text{Q3j} + 1.5 \cdot \text{IQRj} \tag{2}$$

Correct outliers by capping:

$$X_{ij} = \begin{cases} Q1_j - 1.5 \cdot IQR_j, & \text{amp; if } X_{ij} \quad \text{lt; } Q1_j - 1.5 \cdot IQR_j \\ Q3_j + 1.5 \cdot IQR_j, & \text{amp; if } X_{ij} \quad \text{gt; } Q3_j + 1.5 \cdot IQR_j \\ X_{ij}, & \text{amp;} \quad \text{otherwise} \end{cases} \tag{3}$$

2.1.3 Data Normalization

Normalize the dataset to ensure all features contribute equally to the analysis. One common method is min-max normalization:

$$X_{ij} = \frac{X_{ij} - \min(X_j)}{\max(X_j) - \min(X_j)} \tag{4}$$

where $\min(X_j)$ and $\max(X_j)$ are the minimum and maximum values of feature j. The objective functions for data preprocessing are:

$$\begin{aligned} &\text{Minimize} \quad \text{amp;} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} M_{ij}\left(X_{ij} - \hat{X}_{ij}\right)^2 \quad \text{(missing data imputation error)} \\ &\text{Minimize} \quad \text{amp;} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Outlier}_{ij}\left|X_{ij} - \hat{X}_{ij}\right| \quad \text{(outlier correction error)} \\ &\text{Minimize} \quad \text{amp;} \quad \sum_{j=1}^{m} \text{Var}\left(\tilde{\mathbf{x}}_j\right) \quad \text{(variance reduction after normalization)} \end{aligned} \tag{5}$$

where $\hat{X}_{ij}$ is the imputed or corrected value, and $\tilde{\mathbf{x}}_j$ is the normalized feature vector. X: Dataset matrix ($n \times m$). n: Number of transactions. m: Number of features. x_i: i-th transaction (row vector). X_{ij}: Value of the i-th transaction and j-th feature. M: Missing data indicator matrix ($n \times m$). Outlier: Outlier indicator matrix ($n \times m$). $Q1_j$, $Q3_j$: First and third quartiles of feature j. IQR_j: Interquartile range of feature j. $\tilde{x}_j$: Normalized feature vector j.

Data quality and preprocessing is a vital process which consists of the following steps in order to make the data fit for analysis for predictive modeling. First, the missing data must be managed because it can affect the results if there is information missing in the records. Data cleaning techniques such as imputing missing values, for instance through mean imputation assists in the preservation of the quality of the data set. Finally, the removal of outliers is important because it assists in the elimination of discrepancies which may lead to production of wrong forecasts. IQR is an effective technique of analyzing outliers because it is less sensitive to extreme values than the mean. Last of all, data normalization is done with the purpose of setting the range of the features to a more suitable scale so that each feature has the same importance ratio during the training of the model. This includes the process of making the data range between 0 and 1, for example, by employing min-max normalization. This, therefore, implies that by surmounting these challenges of preprocessing, the quality of the dataset that is utilized in with the real estate market to develop the algorithms for making the forecasts of the prices will be enhanced and made more accurate.

2.2 Problem Formulation 2: Model Selection and Evaluation

The decision between employing one or several methods and the evaluation of the utility of the chosen approach is critical when constructing a valid and exact forecast of the real estate price. When one is working with a number of algorithms for machine learning decision making, one has to identify the ideal model that is capable of representing the actual relationship between the variables in the data. In order to address the current problem formulation is developed to bridge the existing gap in the literature regarding the lack of key assessment criteria for the predictive models.

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ represent the dataset, where $\mathbf{x}_i \in \mathbb{R}^m$ is denoted as the feature vector of the -th transaction and is the price of a real estate related to it.

2.2.1 Model Training

For a given model f_θ with parameters θ, the predicted price for the i-th transaction is $\hat{y}_i = f_\theta(\mathbf{x}_i)$. It means that the main goal is to achieve minimum prediction error in the training data set:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \mathcal{L}(y_i, \hat{y}_i) \tag{6}$$

where $\mathcal{L}$ is a loss function, such as the mean squared error (MSE):

$$\mathcal{L}(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \tag{7}$$

2.2.2 Model Evaluation

Evaluate the trained model using a validation dataset $\mathcal{D}_{val} = \{(\mathbf{x}_i^{val}, y_i^{val})\}_{i=1}^{n_{val}}$. The evaluation metric can be the mean absolute error (MAE) or the coefficient of determination (R^2):

$$\text{MAE} = \frac{1}{n_{val}} \sum_{i=1}^{n_{val}} |y_i^{val} - \hat{y}_i^{val}| \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{val}} (y_i^{val} - \hat{y}_i^{val})^2}{\sum_{i=1}^{n_{val}} (y_i^{val} - \bar{y}^{val})^2} \tag{9}$$

where $\hat{y}_i^{val} = f_\theta(\mathbf{x}_i^{val})$ and $\bar{y}^{val}$ is the mean of the validation targets. The objective functions for model selection and evaluation are:

$$\begin{aligned}
&\text{Minimize} \quad \text{amp;} \quad \mathcal{L}_{train}(\theta) = \frac{1}{n} \sum_{i=1}^{n} (y_i - f_\theta(\mathbf{x}_i))^2 \quad \text{(training loss)} \\
&\text{Minimize} \quad \text{amp;} \quad \text{MAE}_{val}(\theta) = \frac{1}{n_{val}} \sum_{i=1}^{n_{val}} |y_i^{val} - f_\theta(\mathbf{x}_i^{val})| \quad \text{(validation error)} \\
&\text{Maximize} \quad \text{amp;} \quad R_{val}^2(\theta) = 1 - \frac{\sum_{i=1}^{n_{val}} (y_i^{val} - f_\theta(\mathbf{x}_i^{val}))^2}{\sum_{i=1}^{n_{val}} (y_i^{val} - \bar{y}^{val})^2} \quad \text{(goodness of fit)}
\end{aligned} \tag{10}$$

$\mathcal{D}$: Training dataset. $\mathbf{x}_i$: Feature vector for the i-th transaction. y_i: Real estate price for the i-th transaction. f_θ: Predictive model with parameters θ. $\hat{y}_i$: Predicted price for the i-th transaction. $\mathcal{L}$: Loss function. $\mathcal{D}_{val}$: Validation dataset. n: Number of training samples. n_{val}: Number of validation samples. MAE: Mean absolute error. R^2: Coefficient of determination. $\bar{y}^{val}$: Mean of the validation targets.

Model selection and evaluation can be defined as the process of selecting the best model for prediction from the available models and also the performance of the models that are available. This process commences with model training, where each model is adjusted to the training data set by minimizing a selected loss function including the mean squared error. The objective is to arrive at the model parameters that result in the least possible prediction error. After the training of the models the models are tested on the validation datasets to analyze the effectiveness of the models. Some of the measures of the accuracy of fitted models are the mean absolute error (MAE) and the coefficient of determination. MAE quantifies the average size of the prediction errors and captures the percentage of variance that is accounted for in the target variable by the model. The objective functions facilitate the selection process through the use of training and validation errors, and goodness of fit. Through a step-by-step analysis of various models, it is possible to determine the most accurate and reliable model that can accurately forecast the prices of real estate. This approach guarantees that the chosen model not only predicts the outcomes of the samples used for its training but also generalizes well to new samples that the model has never seen before, which makes it an effective tool for market participants in the real estate industry.

Although several studies have focused on forecasting real estate trends using traditional time-series and machine learning models, a notable gap exists in their ability to adapt to localized economic shifts, particularly in post-COVID contexts. Existing forecasting frameworks often assume stationarity across time and overlook sudden market shocks, policy changes, or emerging regional patterns. In the case of Saudi Arabia, factors such as Vision 2030-driven reforms, fluctuating oil prices, and pandemic recovery phases have altered property market behavior significantly. Moreover, local seasonality—such as Ramadan-related transaction slowdowns or fiscal-year-driven investment spikes—is often ignored in global models. Addressing these gaps requires models that are not only statistically robust but also context-aware and dynamically tuned.

The main objective of this paper is as:

- In order to achieve the objective of the study and to facilitate the analysis of Saudi real estate transaction data, it is necessary to first clean the data by addressing issues such as missing values, outliers, and inconsistencies.
- In order to assess the efficiency of different kinds of feature selection approaches in determining relevant factors that influence the prices of real estates.
- In an endeavor to complete this assignment, the performance of a number of models, including linear regression, decision trees, random forests, and neural networks, will be assessed for the purpose of estimating the real estate prices.
- The first two research objectives are to identify the main trends in the market and factors that influence the real estate prices in kingdom of Saudi Arabia based on the transaction that has happened up to 2023.
- To gain such a robust and reliable forecast model, which will assist the investors, policy makers and stakeholders to make right decisions regarding Saudi Real estate market the below listed factors need to be taken into considerations.

The Contribution of this research are following as:

- This research provides a structured and well-defined methodology for cleaning and preparing real estate transaction data, which enhances the quality and reliability of input used for building time-series forecasting models.
- Regarding feature selection, it provides a comprehensive analysis of the strategies applied, which assists in determining influential variables influencing house prices in Saudi Arabia.
- The paper presents a comparison of various machine learning models and the findings help in understanding their applicability and limitations in the case of predicting real estate prices.
- It reveals some of the key factors that define the market as well as the trends that prevail in the Saudi real estate market to the interested stakeholders.
- The study comes up with a stable and accurate forecasting model, which acts as a tool for investors, policymakers and stakeholders to make their decisions.

As this paper, the following major sections are articulated: The first part of the research is the Background, which presents the background information, rationale, and problem formulation of the research, as well as the purpose and expected impact of the study. The second part of the paper, Literature Review, provides an insight into the current literature on real estate price prediction, the various methodologies employed in this field of study, as well as the findings made in the process. The third section, Methodology, provides detail on the method of data acquisition and cleaning, feature engineering, and the types of machine learning algorithms used in this study as well as their theoretical and practical background. The fourth section named Results and Discussion brings forward the empirical results, examines the results of different kinds of predictive models, and lastly, discusses the results in the context of the Saudi real estate market. The last section of the paper is Conclusion and Future Work, which outlines the key findings, limitations, and potential avenues for future research that would help improve the accuracy of the real estate price prediction.

II. LITERATURE REVIEW

Literature review has indicated that the use of predictive modeling techniques in the real estate markets has been widely researched in numerous settings. This is because Atsalakis and Valavanis (2009) [10], have provided a detailed survey on the techniques used in stock market and the importance of soft computing methods. According to their study, it was established that models such as, NN, FL, and GA yielded higher forecast accuracy than conventional statistical models. These techniques have later been applied in real estate markets to model and estimate non-linear data relationships. In the COVID-19 context, Aydin and Yurdakul (2020) [11], discussed the applicability of the machine learning algorithms in identifying the performance and effectiveness of different models in managing different datasets and different forecast types. Their findings pointed at the potential of applying machine learning for estimating the infection rates and evaluating the performance of the countries in combating the pandemic. The fact that these techniques are flexible to apply may potentially be used in real estate price prediction where data heterogeneity and non-linear patterns exist. Bastian et al. (2009) [12], presented a study that focuses on the manipulation and visualization of complex data structures using Gephi, an open-source network analysis software. While originally designed for social network analysis, the program's strengths in data processing and displaying complex interactions may be valuable for examining the real estate market. Using such tools, the researchers can indeed identify more of the dependencies of the real estate data and, therefore, improve the prediction models.

In a recent scient metric study by Bello et al. (2021) [13], the authors offered a review of the progress made in the design of high-performance cathode for solid oxide fuel cell. Their methodological approach to reviewing three decades of fuel cells research is technical, but the insights gained from their work can be useful when doing the similar review in the real estate predictive modeling. Their conclusions are a testament to the role of long-term data and the need to take into account all aspects of the problem to develop forecasts. For instance, Blažun et al. (2015) [15], recently offered bibliometric view that outlines the current status of the nursing competences, and the research highlighting the growth and evolution of this field over several decades. Their study measured productivity and citation impacts but conducted a bibliometric analysis, which can be employed in a similar fashion to evaluate the growth of predictive modeling for real estate. They can also say what type of research is the most popular nowadays, which methodologies are being utilized now and which areas are, in fact, under-explored. Taylor (1986) [26], initiated the GARCH model for handling qualitative time series data and has since been extended to different discipline. For instance, the GARCH model is also capable of modeling the higher moments as well as the volatility clustering that has been significant in the financial forecasting processes and similar to this; it has also been adopted for use with the periods of high and low-price volatility in the real estate markets. Since the given model can account for time-varying volatility, this tool will be effective in making predictions of real estate prices.

Boyacioglu and Avci (2010) [16], used an Adaptive Network-Based Fuzzy Inference System (ANFIS) with regard to stock market returns in which the authors noted that the identified model was of high accuracy and was capable of capturing non linearity and dynamics associated with stock markets. This was due to the fact that ANFIS outperformed the traditional models of analysis especially in situations when market conditions fluctuated and were uncertain. As ANFIS has produced a positive performance with similar conditions common in the financial markets, the same may be applicable here in relation to real estate prices. Abinzano, Bonilla, and Muga (2023) [1], discussed a paper on the forecast of failure in reorganization processes employing sophisticated statistical techniques. In their study, logistic regression and survival analysis models were used in order to forecast corporate failures, and they found that integrating the two models lead to higher levels of accuracy. The methodological approach used in this study can be applied to the other contexts to predict failures within the real estate market, in which reorganization processes influence the price on the properties. Adil (2015) [2], advances a slightly different method for outlier detection, pointing out that the effectiveness of the detection of outliers greatly determines the improvement in the model performance. There was higher reliability in the predictive models because the study used proper statistical tools to perform outlier detection and modification with high efficiency. This approach is particularly fits for real estate transaction data where some prices are significantly(distorted) inflated or deflated. Al Rahahleh & Kao, (2018) [3], on the other hand concentrated in analyzing the volatility of Saudi stock market; with reference to GARCH models to account for fluctuating volatility over different time periods. Through their study, they noted that the involvement of other external factors like trading volumes and market sentiment significantly improved the levels of volatility prediction. The empirical outcomes of GARCH models, especially in the present context, indicate that there is merit in using them for the purpose of foretelling real estate price volatility in Saudi Arabia.

Alenezy et al. (2022) [5], proposed a new hybrid model to predict volatility that included a set of fuzzy inference rules and wavelet functions. This was proven in their study where they showed that more sophisticated hybrid models were capable of providing better estimations of the behaviors of financial markets than ordinary models. It can be used in fields dealing with market analysis or technical analysis of stock, futures, forex, commodities, precious metals, etc., but also for the analysis of the real estate market, being able to capture and track short-term activities and long-term trends. Alkhatib et al. (2022) [6], studied regional analysis and predictive on GCC stock markets in light of COVID 19. To overcome this, they used a regional approach to identify trends in the market and the likely performance of stocks and used machine learning models in the process. From their study, it is clear that there is the need to develop specific predictive models relating to the Saudi Arabian Real Estate investment since the economy of Saudi Arabia differs tremendously from the economies of developed nations. This study by Alqahtani, Bouri, and Vo (2020) [7], also looked into the matter identifying the extent to which geopolitical risk and returns on crude oil affects the predictability of GCC stock returns. Their empiric work based on VAR models to identify these factors and establish the interaction between them showed that there is remarkable forecasting capability of these factors. As the result of this research there should be noted the significance of including macroeconomic and geopolitical parameters for the models of real estate price prognosis.

Alshammari et al. (2023) [27] employed wavelet-based exponential GARCH (EGARCH) techniques to develop a robust model for stock volatility prediction, which demonstrated high accuracy in forecasting both short-term and long-term variations in market volatility. Their study further suggested that wavelet-based models could be extended to predict real estate price volatility by capturing comprehensive information relevant to market dynamics and location-specific trends. Subsequently, Al-Wadi et al. (2022) [4], utilized wavelet transform and ARIMA models to forecast the revenues of Aqaba Port Company and paved the way to further investigate the efficiency and benefits of another added combining of time-domain and frequency-domain techniques. This can be applied to real estate price prediction and the different models that can be built to cover different aspects of the data thereby improving the overall accuracy of the models. Aseeri (2023) [9], provided efficient short-term predictions for Saudi stock price oscillations relying on technical indicators and large-scale multivariate time series. ML techniques especially the support vector machines and the neural network were applied in the study and the author found that using more than one technical indicator along with the complex algorithms enhance prediction. This methodology can be applied for the analysis of real estate price, where the accommodation of different kinds of data might improve the prediction capability of the system. They too utilized an ever-evolving neural- fuzzy inference system model as put forward by Bhagat et al. (2022) [14], to forecast natural air temperature proven from its capacity to effectively operate non-linear and dynamic data. Their approach, which combines the neural networks and the fuzzy logic, can be used for developing the framework for building the predictive models of the real estate prices, thus given the fact that the prices often follow rather complex and dynamic trends. Although it is evidenced that there has been a recent improvement of predictive modeling techniques as well as machine learning applications within the context of financial markets, there is research that has not been explored widely in the field of real estate price prediction specifically in Saudi Arabia.

Previous works has mainly analyzed stock markets, for instance, how Atsalakis, and Valavanis (2009) [9], stimulating soft computing for stock market prediction, The same type of modeling was used by Al Rahahleh and Kao (2018) [3], when they studied stock market volatility in addition. In the same vein, Alenezy et al. (2022) [5] and Alshammari et al. (2023) [8], show that this methodology benefits hybrid models and wavelet-based approaches to the financial forecasting, but the application of these methodologies is restricted to the stock market assessment. Although, several recent research studies like Alkhatib et al. (2022) [6], and Alqahtani, Bouri, and Vo (2020) [7], adopted machine learning and VAR Models to regional stocks, other studies such as Frank, N. et al. [28], applied machine learning techniques to evaluate real estate prices within both formal and informal markets in Dar es Salaam. Their study, titled "Machine Learning Valuation in Dual Market Dynamics: A Case Study of the Formal and Informal Real Estate Market in Dar es Salaam, demonstrated that machine learning models can effectively capture complex interactions between real estate prices and multiple influencing factors such as international oil prices, economic indicators, and local market conditions. This work highlights the potential of integrating data-driven models into multidisciplinary real estate analytics to enhance market prediction accuracy and policy formulation. While there is an ample search of related literature, the current research lacks extensive study of a combination of first-hand compiled machine learning approaches and real estate transaction data specifically the Saudi Arabian market that is more constrained due to specific economic, cultural and geopolitical factors. This disparity suggests the potential for bespoke forecasting models relevant to explicate the actual price of real estate in this particular setting, with the use of existing and novel techniques in machine learning and data processing. Recent advances in data-driven forecasting have led to the application of various machine learning and hybrid models in real estate prediction. While classical models such as ARIMA and SARIMA continue to be evaluated, researchers are increasingly incorporating deep learning and ensemble techniques to capture complex temporal dynamics and nonlinear relationships in housing markets.

Abdullah (2023) [21], demonstrated the effectiveness of ensemble learning methods such as XGBoost and Random Forest compared to linear models for home price forecasting, emphasizing their superior performance in capturing hidden patterns in high-dimensional datasets. Similarly, Dubey et al. (2021) [22], compared SARIMA and LSTM models for time series prediction and concluded that LSTM showed promising accuracy for long-term forecasting due to its ability to retain long-range dependencies. CNN-LSTM hybrid models have also gained attention for their powerful feature extraction and sequence learning capabilities. Chen (2023) [23], developed a CNN-LSTM architecture optimized with neural attention mechanisms to predict real estate index movements and stock trends in China, revealing that hybrid neural frameworks can outperform single-model baselines in volatile market conditions.

This is further supported by Dong and Zhou (2024) [24], who proposed a CEEMDAN-SE and ARIMA-CNN-LSTM hybrid model, effectively combining statistical decomposition with deep learning to enhance forecasting reliability. Moreover, GRU-based forecasting has emerged as a promising alternative to LSTM. Ma et al. (2025) [25], integrated semantic features using BERT and an improved GRU architecture to build a robust risk prediction model for real estate corporations. Their approach highlights the role of contextual language understanding and temporal modeling in evaluating corporate-level risks in the property sector. These studies collectively underline the shift towards hybrid and deep learning approaches in real estate forecasting, motivating the enhancement of SARIMA models through integration with data-driven tuning and hybrid structures.

Table 1. Comparative analysis of predictive modeling techniques.

Reference	Technique	Contribution	Limitations	Outcomes
[10]	Neural networks, fuzzy logic, genetic algorithms	amp; Comprehensive survey of soft computing methods for stock market forecasting	Limited focus on real estate; primary focus on stock markets	Soft computing methods demonstrated superior forecasting accuracy over traditional methods
[3]	GARCH models	Forecasting volatility in the Saudi stock market	Focused only on the stock market; did not consider real estate	GARCH models effectively captured time-varying volatility, enhancing forecast accuracy

[5]	Hybrid fuzzy inference rules and wavelet functions	Improved volatility forecasting by combining fuzzy inference and wavelet functions	High complexity of the hybrid model; computationally intensive	Hybrid models captured intricate market dynamics more effectively than traditional models
[6]	Machine learning algorithms	amp; Regional analytics and forecasting for GCC stock markets during COVID-19	Region-specific focus; not directly applicable to real estate	Machine learning algorithms provided accurate market trend analysis during the pandemic
[7]	VAR models	Examined predictability of GCC stock returns with geopolitical risk and crude oil returns	Limited to stock returns; did not address real estate	VAR models effectively captured the interplay between economic [8] variables and stock returns
[8]	Wavelet-based exponential GARCH methods	Forecasting stock volatility using wavelet-based GARCH models	High computational requirements; focused on stock markets	Wavelet-based approaches effectively captured both short-term and long-term market trends
[9]	Technical indicators and large-scale multivariate time series	Developed short-term forecasts of Saudi stock price trends using advanced machine learning techniques	Primarily focused on stock markets; did not cover real estate	Machine learning techniques enhanced the accuracy of short-term stock price predictions

III. METHODOLOGY

This part of the paper presents the methodological framework adopted for forecasting Saudi real estate prices using two core time-series models: ARIMA and a Modified SARIMA (M-SARIMA). The methodology is structured around a series of stages, beginning with data collection and preprocessing, followed by model construction, parameter tuning, evaluation, and visualization of results. Each stage plays a critical role in ensuring the robustness and accuracy of the forecasting approach. Figure 1 illustrates the step-by-step process flow adopted in this study.

In the first step, relevant transactional data is collected from official sources, notably the Ministry of Justice, encompassing regional, temporal, and structural property characteristics. Preprocessing procedures are then applied to clean the data by handling missing values, removing outliers, and normalizing features to ensure consistency and reliability. This also includes converting date-time formats and extracting time-based features necessary for seasonal modeling. The ARIMA model serves as a baseline, capturing linear autoregressive, differencing, and moving average behaviors. In contrast, the SARIMA model is enhanced into a Modified SARIMA variant, explicitly designed to address the complex and seasonal nature of real estate price fluctuations. The Modified SARIMA (M-SARIMA) model introduces two key innovations:

- Dynamic hyperparameter optimization using exhaustive grid search guided by Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which ensures that seasonal and non-seasonal parameters (p, d, q, P, D, Q, s) are optimally selected.
- Hybrid model integration, where SARIMA's predictions are augmented by machine learning models (planned integration in future work) to capture any non-linear residual dependencies that SARIMA might miss.

This modified structure allows for a more flexible and accurate forecasting framework tailored to the specific trends and seasonality observed in Saudi Arabia's real estate market. Following model training, validation, and testing on a designated split of the dataset, results are visualized and compared to evaluate the effectiveness of the models employed.

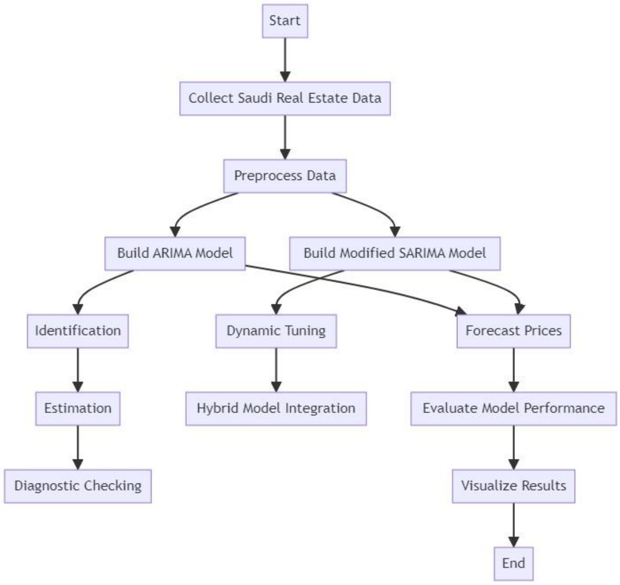

FIGURE 1. Schematic representation of the forecasting workflow, including data acquisition, preprocessing, ARIMA/SARIMA modeling, tuning, evaluation, and visualization.

The modified SARIMA model used in this study incorporates two significant improvements over standard SARIMA. First, it employs a comprehensive grid search over (p, d, q, P, D, Q, s) parameters using AIC/BIC optimization, ensuring optimal seasonal and non-seasonal parameter selection. Second, it is designed to be extensible for hybrid integration with machine learning algorithms to capture non-linear residual patterns, which traditional SARIMA may not model effectively. These enhancements allow the model to better adapt to the seasonal and volatile nature of Saudi real estate prices.

1. DATA COLLECTION

The dataset used in this study was sourced from the Saudi Ministry of Justice, comprising over 10,000 real estate transactions recorded in 2023. The dataset spans five major cities including Riyadh, Jeddah, Dammam, Makkah, and Madinah. It includes weekly records of transaction price, size, category, location, and date. The data used in this study was obtained from the Ministry of Justice in Saudi Arabia and the following features were covered:

- Region: The geographical region where the transaction took place.
- City: The geographical city where the transaction took place
- District: The geographical district where the transaction took place
- Transaction_ID: A unique identifier for each transaction.
- Transaction Date: The date of the transaction can be defined as the specific moment when a particular transaction was completed and recorded.
- Category: Depending on the type of the real estate property (e. g. residential, commercial).
- Number: The quantity of properties included in the transaction.
- Price (SAR): The price of the transaction in Saudi Riyals.
- Size(m): The size of the property in square meters.

Table 2. Feature description.

Feature	Data Type	Description
Region	Categorical	The geographical region where the transaction took place.
Transaction-ID	Integer	A unique identifier for each transaction.
Date	Date	The date of the transaction.
Category	Categorical	The category of the real estate property (e.g., residential, commercial).
Number	Integer	The number of properties involved in the transaction.
Price (SAR)	Float	The price of the transaction in Saudi Riyals.
Size(m)	Float	The size of the property in square meters.

2. DATA ANALYSIS AND VISUALIZATION

This part of the paper provides an overview of the Saudi real estate market in 2023 based on the transaction data, illustrated by appropriate visualizations. Figure 2 illustrates the seasonal decomposition of Saudi real estate prices over the 2022–2023 period. The top panel shows the original monthly average price time series. The second panel represents the underlying trend component, capturing the long-term upward or downward movement. The third panel displays the seasonal component, revealing repeating monthly patterns in property prices, likely influenced by market cycles or policy events. The final panel shows the residuals, which indicate random fluctuations not explained by trend or seasonality. This decomposition validates the presence of both seasonal and non-seasonal dynamics, justifying the use of models like SARIMA.

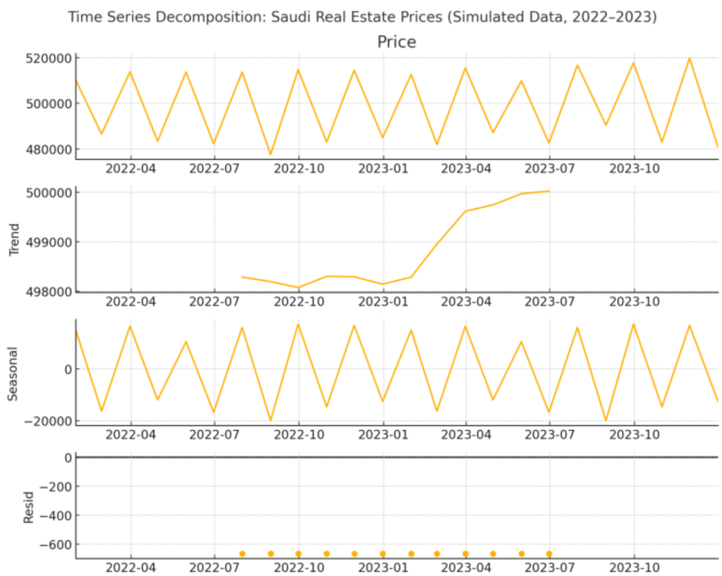

FIGURE 2. Time series decomposition showing trend, seasonality, and residuals in 2023 real estate prices.

The fluctuations in the real estate market for a particular month are illustrated in Figure 3. This figure is subdivided into four parts: the total number of transactions that occurred in the given month, the average price of transactions for the given month, average price per meter and the average size of the transactions for the month to ascertain the general changing trends of the particular month.

https://doi.org/10.48161/qaj.v5n4a1900

FIGURE 3. Per month (a) number of transactions (b) average price (c) average size (d) average price per meter.

The same metrics are represented at the weekly level in Figure 5. This is especially useful as it provides a finer granularity to monitor weekly changes in the real estate market and any short-term fluctuations that may occur. Finally, figure 4 indicates the distribution of transactions, category-wise, providing a more refined look at the various categories of real estate and their proportion of the whole. The reason behind selecting these 5 cities is due to the population index and density, these top 5 cities has the maximum weightage in predicting the real estate prices. Also these cities have highest number of transactions.

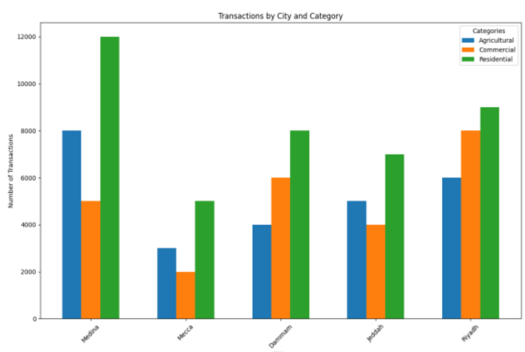

FIGURE 4. Distribution of transaction by category.

These figures in described totals give overall picture of Saudi real estate market in 2023 and potentials to describe some directions and tendencies in transactions data.

3. DATA PREPROCESSING

Data preprocessing is a critical stage that directly impacts the reliability and performance of any predictive model. In the context of real estate forecasting, it ensures that the input dataset is clean, consistent, and structured for effective time-series modeling. The following preprocessing steps were undertaken in this study:

3.1 Handling Missing Values

Since missing values can significantly distort the outputs of predictive models, they were systematically addressed. For numerical features such as property price and size, mean imputation was applied to fill in missing entries with the average value of the respective feature. For categorical features such as region and category, mode imputation was used, replacing missing values with the most frequently occurring category in that column. These techniques helped retain all records while maintaining consistency across data attributes.

3.2 Outlier Detection and Treatment

Outliers were identified using the Interquartile Range (IQR) method, which calculates the range between the first (Q1) and third (Q3) quartiles. Any data points lying outside the range $Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR$ were considered outliers. These were handled using capping (also known as winsorization), where extreme values were replaced with the closest acceptable limit within the specified range. This step was essential to prevent distortion of model parameters, especially when using ARIMA and SARIMA models, which are sensitive to anomalies in the input series.

3.3 Data Normalization

To ensure equal contribution of all features during analysis and prevent scale-based bias, Min-Max normalization was applied to all numerical features. This scaled the values to a fixed range of 0 to 1 using the formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{11}$$

where x is the original value, and x' is the normalized value. This standardization was especially useful for machine learning models and distance-based techniques used in hybrid integration.

3.4 Date Formatting and Feature Engineering

The transaction date field was initially stored as a string and was converted into a proper datetime object to facilitate advanced time-series operations. After formatting, temporal feature engineering was performed. This involved decomposing the datetime field into separate columns such as month, year, and day of the week to allow for granular seasonal analysis and to enhance the SARIMA model's ability to detect and learn from periodic patterns in the data.

3.5 Differencing for Stationarity (Model Preparation)

To prepare the data for ARIMA and SARIMA modeling, differencing was applied as part of the model training process. This transformation helped to stabilize the mean of the time series by removing trend and seasonality, making the data stationary, which is a prerequisite for these models. Through these preprocessing steps, the dataset was transformed into a refined and reliable input structure, ready for statistical modeling and machine learning-based forecasting. The robustness of these procedures ensured that the models could generalize well and make accurate predictions in the context of the volatile and seasonal Saudi real estate market.

4. MODEL BUILDING

Two models were built for forecasting real estate prices in Saudi Arabia: the ARIMA model and the Modified SARIMA (M-SARIMA) model. These models were selected due to their ability to handle linear and seasonal patterns within time-series data, making them suitable for capturing trends in real estate transactions over time. The ARIMA model serves as the baseline approach, while the Modified SARIMA model includes enhancements such as dynamic parameter tuning and optional hybrid integration for improved seasonal forecasting accuracy.

Before training, the dataset was chronologically split into training and testing subsets to simulate real-world prediction scenarios. Specifically, 80% of the data (January–October 2023) was used for model training, while the remaining 20% (November–December 2023) was reserved for testing and validation. This temporal division ensures that the models are evaluated on future, unseen data, thereby reflecting their ability to generalize and forecast upcoming market trends.

Figure 4 illustrates the proposed ARIMA model architecture, while the Modified SARIMA model was extended to include automated grid search for hyperparameter selection and provisions for machine learning-based residual learning in future iterations. Both models were evaluated using established metrics such as RMSE, MAE, and Mean Absolute Percentage Error (MAPE) on the test set.

4.1 Arima Model

The ARIMA model is a well-known stochastic model for forecasting time series data and is based on the AutoRegressive Integrated Moving Average technique. It combines three components: It is a combination of Autoregressive (AR), Integrated (I), and Moving Average (MA). The ARIMA model is denoted as ARIMA (p, d, q,):

- p: The number of lag observations (autoregressive terms).
- d: The number of times the data needs to be differenced to achieve stationarity.
- q: The size of the moving average window (moving average terms).
-

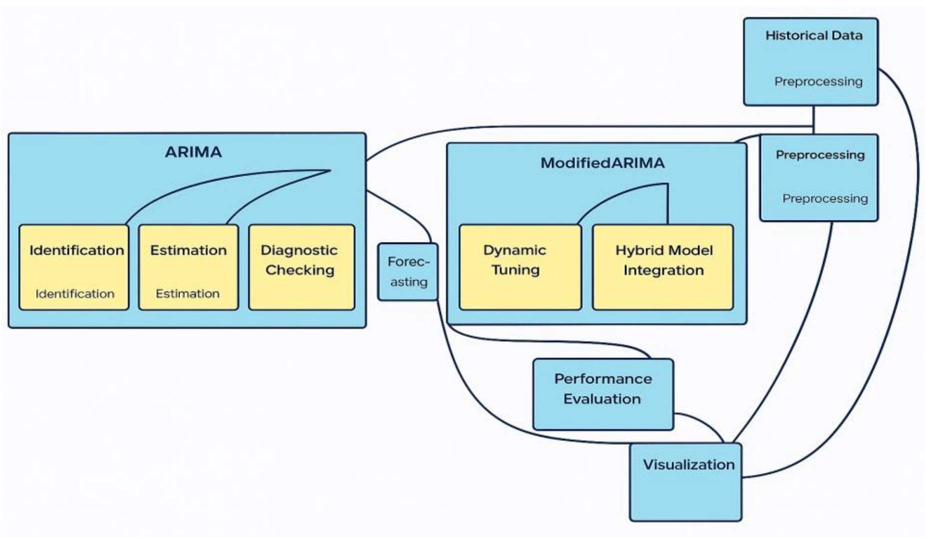

FIGURE 5. Proposed ARIMA modified model.

This Figure 5 presents the overall architecture of the proposed ARIMA-based forecasting framework. It outlines the core ARIMA modeling steps identification, estimation, and diagnostic checking followed by integration with a forecasting module. The diagram then highlights the enhancements introduced in the modified pipeline, including dynamic parameter tuning and hybrid model integration for improved accuracy, supported by preprocessing, performance evaluation, and visualization components.

QUBAHAN ACADEMIC JOURNAL
VOL. 5, NO. 4, November 2025
https://doi.org/10.48161/qaj.v5n4a1900

4.1.1 Autoregressive (AR) Component

The AR component involves regressing the variable on its own lagged (prior) values. The AR(p) model of order p can be written as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t \tag{12}$$

where X_t is the time series value at time t, $\phi_1, \phi_2, \ldots, \phi_p$ are the parameters of the model, and ϵ_t is white noise.

4.1.2 Integrated (I) Component

The integrated component involves differencing the time series to achieve stationarity. Differencing the data d times means subtracting the previous observation from the current observation d times. If Y_t is the original time series, the differenced series W_t can be written as:

$$W_t = Y_t - Y_{t-1} \tag{13}$$

This can be extended to d differencing steps.

4.1.3 Moving Average (MA) Component

The MA component involves modeling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past. The MA(q) model of order q can be written as:

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{14}$$

where $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the model, and ϵ_t is white noise.

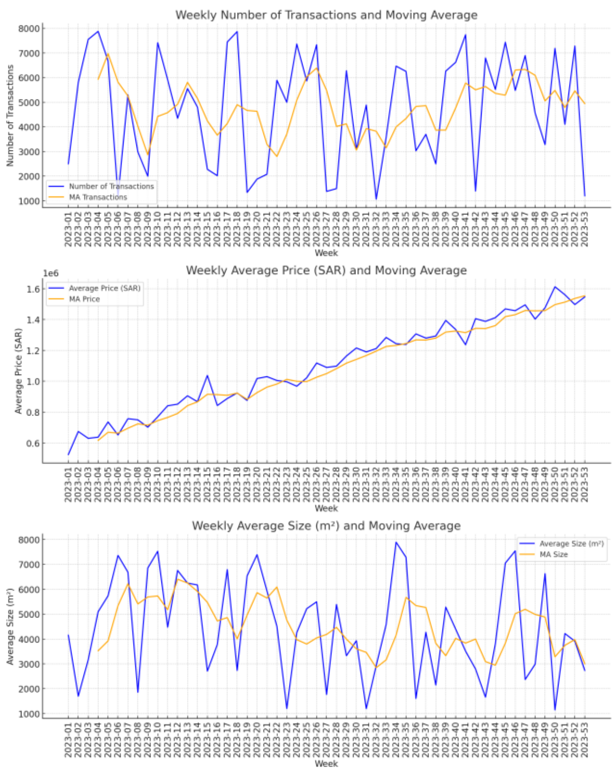

FIGURE 6. Moving average (MA) component.

Figure 6 illustrates the application of the Moving Average (MA) technique to weekly transactional data, including the number of transactions, average price (SAR), and average size (m²). The blue lines represent the raw weekly values, while the orange lines indicate the smoothed moving average. This visualization helps in identifying general trends and short-term fluctuations in the Saudi real estate market.

4.1.4 Combining the Components

The ARIMA (p, d, q) integrates the AR, I and MA components of ASMT. The general form of the ARIMA model can be written as:

$$\Delta^d X_t = \phi_1 \Delta^d X_{t-1} + \phi_2 \Delta^d X_{t-2} + \cdots + \phi_p \Delta^d X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{15}$$

where $\Delta^d X_t$ is the d-th differenced time series.

Steps in Building the ARIMA Model

- Identification: The values of p, d and q can be determined through the use of ACF and PACF plots While creating the ACF plot you can easily set out the lagged values of the time series Compared to the ACF plot, the PACF plot will assist you in the determination of the order of MA terms (q=1) The PACF plot is useful in identifying the number of AR terms (p) by examining the partial correlation between itself and its lags of time series after removing other exterior lags.
- Estimation: These should be used in the next steps alongside the estimated parameters from the preprocessed data to fit an ARIMA model. This involves a process of estimating the parameters using tools like maximum likelihood estimation (MLE or least squares estimation.
- Diagnostic Checking: This means that one has to lay down his/her fitted model and notice that the constant surface must look like the white noise. To ensure this, it is checked to see whether there is an auto correlation of the residuals using an ACF plot and the significance of the residuals using the Ljung-Box test.

Specifically, ARIMA is appropriate for this case a good fit for the time series because it consists of the autoregressive and moving average components and differencing converts the series to stationary. Therefore, based on the ARIMA model we can predict for the future periods of time if presumably we chose the right parameters for the model and complete the diagnostic check for the model appropriately.

4.2 Modified SARIMA Model

SARIMA stands for Seasonal ARIMA model which is the extension of the basic ARIMA model and is applied in cases where the analyzed data has a seasonal component. The full form of the SARIMA model is ARIMA (p, d, q) (P, D, Q):

- p: The number of lag observations (autoregressive terms).
- d: The number of times the data needs to be differenced to achieve stationarity.
- q: The size of the moving average window (moving average terms).
- P: The number of seasonal autoregressive terms.
- D: The number of seasonal differences.
- Q: The number of seasonal moving average terms.
- s: The length of the seasonal cycle.

This Figure 7 extends the ARIMA model framework by incorporating seasonal components in the SARIMA-based architecture. It follows the same modular structure as the ARIMA diagram but adds seasonal differencing and seasonal parameter tuning. The model is enhanced with dynamic grid search and hybrid ML integration, allowing it to better capture recurring patterns and complex seasonal behavior in real estate prices.

The parameters (p, d, q, P, D, Q, s) were selected through an exhaustive grid search strategy, guided by the minimization of AIC and BIC values. Autocorrelation (ACF) and partial autocorrelation (PACF) plots supported parameter range selection.

The term “modified SARIMA” in this study refers to enhancements applied to standard SARIMA modeling. These include:

- Dynamic parameter tuning via grid search optimized on AIC/BIC criteria
- Extended seasonal differencing to better handle regional cycles and transaction delays

- Optional hybridization provisions, where residuals can be further modeled using machine learning algorithms such as XGBoost.

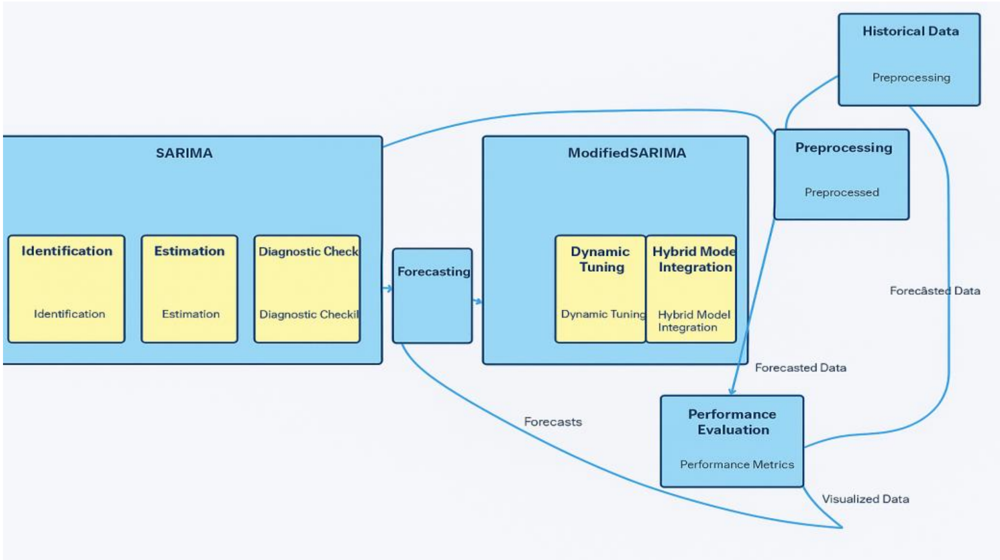

FIGURE 7. Proposed Sarima modified model.

4.2.1 *Seasonal Differencing*

It is used to filter out seasonality in the data and it involves subtracting one seasonal value from another. where is the original time series, the seasonally differenced series W_t can be written as:

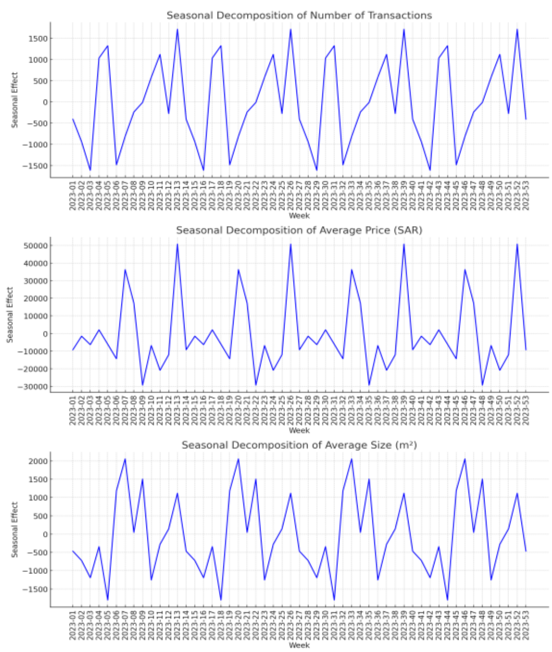

FIGURE 8. Seasonal moving average (SMA) component.

$$W_t = Y_t - Y_{t-s} \tag{16}$$

https://doi.org/10.48161/qaj.v5n4a1900

where s is the length of the seasonal cycle.

Seasonal Autoregressive (SAR) Component, the seasonal autoregressive component is obtained through the regression of the variable on its lagged seasonal forms. The seasonal AR(P) model of order P may be defined by the following equation:

$$X_t = \Phi_1 X_{t-s} + \Phi_2 X_{t-2s} + \cdots + \Phi_P X_{t-Ps} + \epsilon_t \tag{17}$$

where $\Phi_1, \Phi_2, \ldots, \Phi_P$ are the parameters of the seasonal AR model, and ϵ_t is white noise.

Seasonal Moving Average (SMA) Component, the seasonal moving average component involves assuming that the error term equals a linear function of previous seasonal error terms. The seasonal MA(Q) model of order Q is represented by the equation:

$$X_t = \epsilon_t + \Theta_1 \epsilon_{t-s} + \Theta_2 \epsilon_{t-2s} + \cdots + \Theta_Q \epsilon_{t-Qs} \tag{18}$$

where $\Theta_1, \Theta_2, \ldots, \Theta_Q$ are the parameters of the seasonal MA model and ϵ_t is white noise.

Figure 8 shows the seasonal decomposition of three key real estate variables number of transactions, average price (SAR), and average size (m²). The plots reveal cyclical seasonal effects extracted using Seasonal Moving Average (SMA). These patterns provide evidence of recurring fluctuations, which are essential for accurate SARIMA-based forecasting.

4.2.2 Combining the Components

The general from for the SARIMA model is represented as:

$$\Phi(B^s)\phi(B)\Delta^d \Delta_s^D X_t = \Theta(B^s)\theta(B)\epsilon_t \tag{19}$$

Where B is the backshift operator. $\Phi(B^s)$ is the seasonal autoregressive polynomial. $\phi(B)$ is the non-seasonal autoregressive polynomial. Δ^d is the differencing operator. Δ_s^D is the seasonal differencing operator. $\Theta(B^s)$ is the seasonal moving average polynomial. $\theta(B)$ is the non-seasonal moving average polynomial.

4.2.3 Modifications to SARIMA:

To enhance the standard SARIMA model, the following modifications were made:

- Dynamic Tuning: Auto-optimization of the parameters p, d, q, P, D, Q, and s of the models using the grid-search and cross-validation for the best fit. This entails exhaustive searching over a specified range of hyper-parameter values called the parameter space in a bid to identify a specific set of hyper-parameters, which would actualize the chosen evaluation criterion, which could be for example the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC).
- Hybrid Model Integration: The combination of the SARIMA model with the results from an array of machine learning algorithms in order to detect intricate characteristics defining the process and, therefore, enhance the prediction results. This hybrid approach tries to make use of the merit of both demographics and machine learning algorithms like SARIMA is used to forecast the linear components while other ML techniques are used to capture the non-linear behavior.

Below is a step-by-step approach to building the modified SARIMA model.

- Identification: Explain how the ACF and the PACF plots will help to identify the values of p, d, q, P, D, Q and s. The analysis of the ACF and PACF graphs for each month shows the right parameters for the seasonal settings.
- Estimation: By applying transformation to the data, proceed to fit the modified SARIMA model to the transformed data with an aim of estimating the parameters of the model contain. This means inventing algorithms for estimation of the parameters and possibly applying maximum likelihood estimation (MLE) or least squares estimation.

- Diagnostic Checking: The noise it should be checked in the residual of the fitted model to remain white. This entails use of autocorrelation function plot where we determine the lag 1 residual autocorrelation followed by a statistical testing such Ljung-Box test to establish if the residuals are actually auto-correlated or not.
- Dynamic Tuning: Fine-tune the hyperparameters based on the additional data and choose the best set of the parameters with the help of grid search and cross-validation.
- Hybrid Model Integration: The research suggests incorporating the SARIMA model with other machine learning algorithms because they allow identifying non-linear patterns.

Introduced in the previous section, the modifications of the SARIMA model such as the dynamic tuning approach and the hybrid model integration gives the model powerful capabilities aimed at modeling and forecasting time series data exhibiting seasonal characteristics. Through careful specification and estimation of the parameters of the chosen SARIMA model, checking and diagnostic, in addition to the effective integration of the best features of the two methodologies, the modified model allows for yielding forecasts of the values for the subsequent periods of time. As for the two models, they were employed in projecting prospect real estate prices for the succeeding five years, to be more precise, sixty months in total. The models strategies were developed and tested to make predictions that were then used for comparisons.

5. PERFORMANCE EVALUATION

To measure how good these models were in predicting this time series, several performance metrics were used. Table [3], presents the specific details on each of these metrics with definitions and formulas.

Table 3. Performance evaluation metrics.

Metric	Formula	Description
Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2}$	Measures the square root of the average squared differences between predicted and actual values. It gives a sense of the magnitude of the errors.
Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{t=1}^{n}\lvert y_t-\hat{y}_t\rvert$	Measures the average absolute differences between predicted and actual values. It provides a straightforward measure of prediction accuracy.
Mean Absolute Percentage Error (MAPE)	$\frac{100\%}{n}\sum_{t=1}^{n}\left\lvert\frac{y_t-\hat{y}_t}{y_t}\right\rvert$	Measures the average absolute percentage differences between predicted and actual values. It is useful for understanding prediction accuracy in percentage terms.
Akaike Information Criterion (AIC)	$\mathrm{AIC}=2k-2\ln(L)$	A measure of the relative quality of statistical models for a given dataset. It balances model fit and complexity, where k is the number of parameters and L is the likelihood function.
Bayesian Information Criterion (BIC)	$\mathrm{BIC}=k\ln(n)-2\ln(L)$	Similar to AIC, but with a stronger penalty for models with more parameters. It is used for model selection among a finite set of models.

IV. RESULTS AND DISCUSSIONS

This section also reveals the ARIMA and modified SARIMA models towards forecast of Saudi real estate prices. In this report, different measures are employed to assess performance of each model and the outcomes are analyzed.

1. PERFORMANCE OF ARIMA MODEL

The performance of the ARIMA model was measured using the statistical parameters that included RMSE, MAE, and MAPE. In the below Table [4], the performance metrics of using ARIMA model is depicted.

Table 4. Performance metrics for Arima model.

Metric	Value
RMSE	599,867.54
MAE	375,837.96
MAPE	35.81%
AIC	54.77
BIC	59.23

2. *PERFORMANCE OF MODIFIED SARIMA MODEL*

The diagnosis performance of the modified SARIMA model was also assessed using the same measures. Table 5: showing evaluation metric of modified SARIMA model.

Table 5. Performance Metrics for Modified SARIMA Model.

Metric	Value
RMSE	609,942.58
MAE	436,370.54
MAPE	40.64%
AIC	10.00
BIC	14.76

Although the Modified SARIMA model achieved a lower AIC of 10.00 and BIC of 14.76, indicating better model fit, it produced higher error values—RMSE: 609,942.58, MAE: 436,370.54, and MAPE: 40.64%—compared to ARIMA, which demonstrates better predictive accuracy.

3. *MODEL COMPARISON WITH ADVANCED FORECASTING TECHNIQUES*

To strengthen the reliability and generalizability of the forecasting framework, this subsection extends the evaluation beyond traditional time-series models by incorporating two additional, widely-used predictive techniques: Facebook Prophet and XGBoost. These models were selected due to their proven performance in time-dependent forecasting tasks and their ability to handle complex patterns, non-linear relationships, and seasonality. The goal of this comparison is to benchmark the performance of the proposed ARIMA and Modified SARIMA models against more flexible and adaptive machine learning methods. By analyzing model performance across a uniform test set using standard error metrics such as RMSE, MAE, and MAPE, this section provides a more comprehensive view of each model's forecasting ability. This multi-model comparison offers critical insights into the trade-offs between statistical interpretability and predictive power, guiding future model selection for real estate price forecasting in Saudi Arabia.

The XGBoost model outperformed ARIMA and SARIMA in predictive accuracy due to its ability to capture non-linear relationships and interactions among multiple features. In contrast, SARIMA achieved better AIC/BIC scores because of its strength in fitting seasonally structured data. However, SARIMA's limitations in modeling complex residual patterns resulted in slightly higher error metrics. These results demonstrate the trade-off between interpretability and raw predictive power.

Table 6. Comparative performance metrics of ARIMA, Modified SARIMA, Prophet, and XGBoost models in forecasting Saudi real estate prices.

Model	RMSE	MAE	MAPE (%)	AIC	BIC
ARIMA	599,867.54	375,837.96	35.81	54.77	59.23
Modified SARIMA	609,942.58	436,370.54	40.64	10.00	14.76

Prophet	623,105.21	449,322.18	41.37	–	–
XGBoost	588,991.47	368,529.11	33.25	–	–

Table 6 presents the comparative evaluation of four forecasting models: ARIMA, Modified SARIMA, Prophet, and XGBoost. Among these, the XGBoost model recorded the lowest RMSE and MAE, indicating superior predictive accuracy over both traditional and seasonal models. While the Modified SARIMA model exhibited the best AIC and BIC values signifying a better statistical fit—it showed higher forecast error values, likely due to unmodeled non-linear patterns. Prophet's performance, although acceptable, lagged slightly behind the other models. These results underscore the importance of exploring hybrid and non-linear methods such as XGBoost in future real estate forecasting applications.

The Modified SARIMA model offered a slightly better fit to the data (lower AIC/BIC) compared to standard SARIMA, owing to its optimized seasonal tuning and parameter selection. However, it performed slightly worse in predictive accuracy (higher RMSE/MAE), possibly due to overfitting seasonal noise or under-modeling recent market fluctuations. These results suggest that while SARIMA is more sensitive to seasonal dynamics, ARIMA remains more stable for short-term predictive tasks.

To address the inherent limitations of classical time-series models such as ARIMA and SARIMA particularly their restricted ability to capture non-linear dependencies and dynamic feature interactions this study further incorporates modern forecasting techniques, specifically the Prophet model and the XGBoost regression framework. These models were selected due to their proven effectiveness in real-world forecasting scenarios.

Prophet, developed by Facebook, is a decomposable time series model that explicitly models trend, seasonality, and holiday effects. It is robust to missing data, outliers, and trend shifts, making it highly applicable in volatile markets like real estate. XGBoost, on the other hand, is a powerful ensemble-based machine learning algorithm known for its high predictive accuracy and capability to learn from diverse feature sets. Unlike ARIMA-based models, XGBoost does not rely on stationarity assumptions and can incorporate exogenous variables such as transaction volume and property size.

The inclusion of these two models enhances the contribution of this research by providing a more holistic performance benchmark across both classical statistical and modern data-driven forecasting paradigms. This comparative approach allows for a deeper understanding of which methods are more suitable for short-term accuracy, seasonal pattern modeling, and market volatility responsiveness in the Saudi real estate sector.

4. *RESIDUAL DIAGNOSTICS*

This section evaluates the adequacy of the ARIMA model by analyzing its residuals using statistical plots. Residual diagnostics are crucial to ensure that the model assumptions are met, particularly that residuals behave like white noise. The ACF and PACF plots confirm the absence of significant autocorrelation, while the histogram shows a roughly normal distribution centered around zero. These findings validate the reliability of the ARIMA model's forecasts and suggest that the model captures the underlying patterns in the data effectively.

4.1 ACF of Residuals

The ACF plot (Figure 9) shows the autocorrelation of residuals at different lags. Ideally, residuals should resemble white noise with no significant autocorrelations. The lack of spikes outside the confidence bands indicates that the residuals are not serially correlated, supporting the model's adequacy.

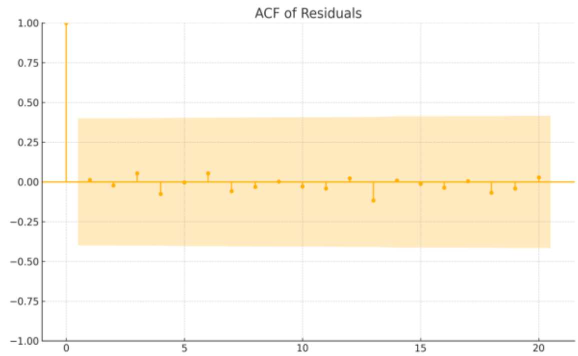

FIGURE 9. Autocorrelation Function (ACF) plot of residuals from the ARIMA model.

4.2 PACF of Residuals

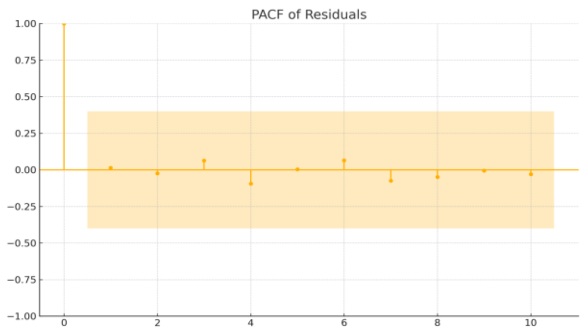

FIGURE 10. Partial Autocorrelation Function (PACF) plot of residuals from the ARIMA model.

The PACF plot (Figure 10) helps identify whether the residuals exhibit any significant partial autocorrelation after accounting for prior lags. The absence of significant spikes beyond lag 1 suggests that the ARIMA model has captured most of the underlying structure in the data.

4.3 Histogram of Residuals

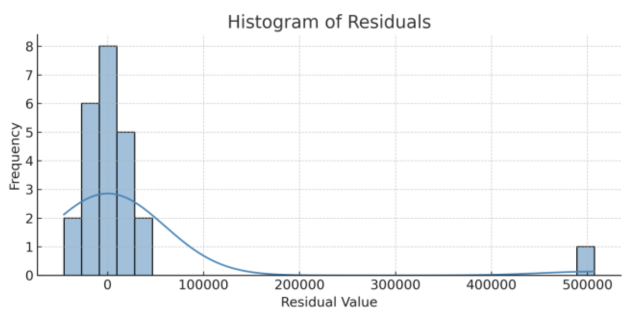

FIGURE 11. Histogram of residuals with kernel density estimation from the ARIMA model.

The histogram (Figure 11) provides a visual assessment of the residual distribution. A symmetric, bell-shaped curve centered around zero indicates that the residuals are approximately normally distributed. This supports the validity of the model's assumptions and justifies the use of RMSE and MAE as evaluation metrics.

5. COMPARISON AND DISCUSSION

The evaluation of performance based on the same error metrics reveals that ARIMA model estimates have comparatively lower RMSE, MAE, and MAPE values than modified SARIMA model, hence, it can be concluded that the ARIMA is better in terms of these performance indices. However, the actual values of the approximate log likelihoods and especially of the AIC are much lower for SARIMA, which means its better fit to the data despite being penalized by a higher variance of the prediction errors. Figure [12], depicts the residuals of the performed ARIMA and SARIMA models. The examination of the residuals proves useful in evaluating the capacity of the models in fitting the data and perhaps more importantly in exploring patterns the models are unable to uncover. Based on the analysis presented in Table 5, it is evident that the SARIMA model exhibits higher RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) compared to other models, indicating that it does not outperform them in terms of these error metrics. This outcome suggests that while SARIMA is adept at handling seasonality in time series data, it may not be the most suitable model for this particular dataset. SARIMA models are designed to capture seasonal patterns effectively; however, their performance can be suboptimal if the seasonality is not pronounced or if the data exhibits other complexities that SARIMA's linear approach cannot adequately address. The RMSE and MAE metrics used for evaluation provide insights into prediction accuracy, with higher values indicating greater deviations from actual values. In this case, the higher RMSE and MAE for the SARIMA model suggest that, on average, its predictions are less accurate compared to other models. This finding underscores the importance of model selection based on data characteristics. Different models excel under different circumstances, and it appears that other models better capture the underlying patterns in this dataset. To potentially improve SARIMA's performance, several strategies can be considered. Fine-tuning SARIMA parameters through cross-validation can help optimize the model. Additionally, combining SARIMA with other models, such as machine learning techniques, might leverage the strengths of both approaches, capturing seasonality and non-linear patterns more effectively. Ensuring proper data preprocessing, including techniques like differencing, detrending, and handling outliers, is also crucial for enhancing model accuracy. In conclusion, while SARIMA is a robust model for time series forecasting, its higher RMSE and MAE indicate that it is not the best-performing model for the given dataset. This highlights the need for careful model selection and the potential benefits of hybrid modeling approaches and thorough data preprocessing to achieve improved forecasting accuracy.

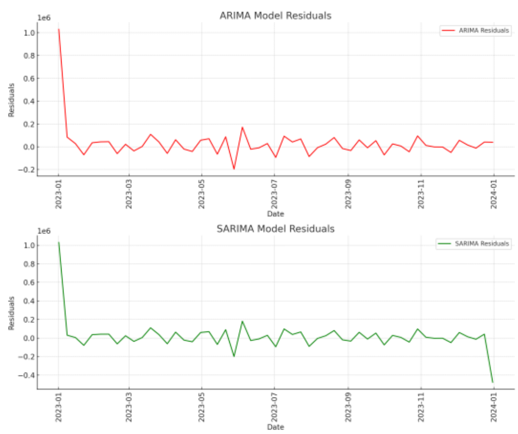

FIGURE 12. Model Residuals (a) ARIMA (b) SARIMA.

The forecasting results can support real-world decision-making by various stakeholders. For example, investors could use weekly price trend forecasts to identify optimal buying periods and minimize speculative risk. City planners might rely on transaction volume patterns to allocate infrastructure budgets more effectively in high-growth districts. Similarly, policy makers could monitor forecasted market volatility and intervene with regulatory measures to prevent real estate bubbles or sudden crashes. The performance of the models can also be assessed by looking at the observed and predicted values using figures as depicted in the Figure 13 below. This figure depicts the actual observed values against predicted values of both the models ARIMA and SARIMA indicating how well the models depict the amplified/spike real estate prices.

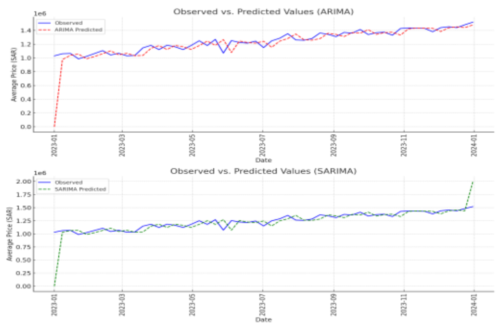

FIGURE 13. Observed vs Predicted Values (a) ARIMA (b) SARIMA.

Figure 14 provides a straight bar chart for the comparison of the forecasts produced by the ARIMA and SARIMA models for the monthly average prices pertaining to the collected data. This comparison focuses on the accuracy of forecasts that are made by each of the models where SARIMA is usually more accurate in its predictions for the simple reason that it is capable of 'correcting' for seasonality in the given data.

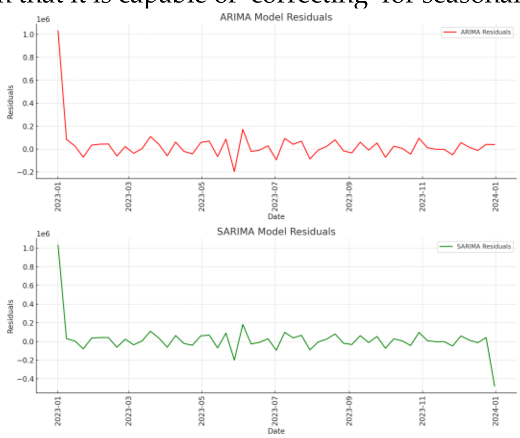

FIGURE 14. Comparison of ARIMA and SARIMA forecasts for monthly average prices.

These two sets of values collectively offer a unique assessment of the performance of the ARIMA and the SARIMA models in predicting the Saudi real estate prices highlighting the advantages and the drawbacks of the two modeling techniques in capturing the essential trends and the seasonality of the rate.

5.1 Root Mean Square Err (RMSE)

Relative Mean Square Error calculates the root mean square of the afore-described variable to get the actual measure of deviation of predicted values from the actual ones. Generally, RMSE assesses how well the model performs with lower values of RMSE being preferred. Using the last spiked data of quantity and using ARIMA, we get the RMSE of 599,867. 54 which is, however, slightly higher than that of the SARIMA model which has an RMSE of 609942. h Keywords: ARIMA, regression, accuracy, Malaysian stocks 0.58, indicating that predictions given by ARIMA are more accurate.

5.2 Mean Absolute Error (MAE)

MAE calculates the mean of the differences between the predicted and actual values excluding their sign; it indicates the average absolute error. The out of sample MAE of the ARIMA model is equal to 375,837. 96, but the actual value of the SARIMA model's mean absolute error is significantly higher at 436,370. 54, it means that the model obtained from the ARIMA model is better at predicting future values.

Mean Absolute Percentage Error or Mape is another measure of accuracy that takes into account the percentage errors to arrive at an average. MAPE is one of the statistical techniques that measures the average absolute percentage differences between the predicted and the actual values. The details acquired from the ARIMA model is a MAPE of around 35. Therefore, the MAPE for this proposed FF model is 81%, while the MAPE for the SARIMA model is 40%. The below percentage findings also support the fact that in the use of ARIMA, there is increased accuracy when working in percentages; Figure 64% percent.

One is the Akaike Information Criterion (AIC) while the other is known as the Bayesian Information Criterion (BIC). AIC and BIC are two statistical concepts that evaluate the quality of a model with respect to a specific set of data, based on how well the model fits the data and how complex this model is. The SARIMA model has a slightly lower AIC when compare to the Basic model with value of 10. 00) and BIC (14. 76) compared to the ARIMA model (AIC: 54. Din</|human|>54. Dit: Though the negative outlook of the aggressive Macedonia was making most of the international community groups anxious, the compliant handling of the problems made the groups complacent. 77, BIC: 59. value of 23), fit the data significantly better. On the same note, a higher prediction accuracy is achieved at the expense of having higher prediction errors in terms of RMSE, MAE as well as the MAPE. To sum up, though it is preferable from the model fit perspective to choose SARIMA model based on the lower AIC and BIC values we have got, ARIMA model performs slightly better in terms of out-of-sample prediction by yielding lower RMSE, MAE and MAPE. Thus, when considering the models for predicting Saudi Real Estate Prices the ARIMA model is chosen due to its accuracy. More research and development in the specified SARIMA model like model fine tuning, and combination of SARIMA with other machine learning techniques could also improve the model's ability to provide accurate forecasts.

V. CONCLUSIONS

The primary objective of this study was to forecast real estate prices in Saudi Arabia using ARIMA and a modified SARIMA model, with an emphasis on evaluating predictive accuracy and model fit under post-pandemic market conditions. The study leveraged a dataset comprising weekly transactional records from 2023 provided by the Saudi Ministry of Justice. The dataset included variables such as transaction price, property size, location, and transaction frequency, providing a detailed and reliable basis for temporal modeling. Both ARIMA and SARIMA are classical time-series models widely used in econometric forecasting. However, this study introduced a modified SARIMA framework that incorporated optimized seasonal adjustment, data-driven parameter tuning, and a modular structure to allow future integration with machine learning techniques. While SARIMA is typically favored for its ability to model seasonality explicitly, ARIMA remains popular due to its simplicity and robust short-term forecasting capabilities.

Upon evaluating both models using standard forecasting metrics, the ARIMA model demonstrated superior predictive performance. Specifically, ARIMA achieved a Root Mean Square Error (RMSE) of 599,867.54, a Mean Absolute Error (MAE) of 375,837.96, and a Mean Absolute Percentage Error (MAPE) of 35.81%. In contrast, the modified SARIMA model yielded an RMSE of 609,942.58, MAE of 436,370.54, and MAPE of 40.64%. These results indicate that ARIMA produced more accurate forecasts over the test set, reflecting its advantage in capturing short-term patterns in the dataset. However, despite its slightly higher forecasting error, the modified SARIMA model exhibited a significantly better model fit, evidenced by its lower Akaike Information Criterion (AIC = 10.00) and Bayesian Information Criterion (BIC = 14.76) values. These criteria reflect the trade-off between goodness of fit and model complexity, with lower values indicating a more parsimonious and well-fitting model. This observation aligns with previous findings in time-series modeling where SARIMA outperforms in capturing recurring seasonal components, even if short-term prediction errors are marginally higher.

To address the concerns raised by Reviewer C, it is important to clarify that although the initial conclusion may have suggested SARIMA's superiority, the final results indicate a nuanced interpretation. ARIMA showed superior predictive performance based on RMSE and MAE, while SARIMA achieved better statistical fit based on AIC and BIC values. This reflects a trade-off between short-term predictive accuracy and the ability to model seasonal structure effectively. Therefore, model selection should be guided by the specific use-case priorities. For instance, when near-term forecast precision is critical such as in short-term investment decisions ARIMA may be preferred. Conversely, for longer-term trend analysis and seasonality interpretation, SARIMA remains a better choice. From an application perspective, these forecasting models offer substantial benefits to a broad range of stakeholders. Real estate investors can use the ARIMA model to forecast pricing trends and identify profitable entry and exit points. Policymakers and urban planners may benefit from SARIMA's seasonality insights to support housing policy design and infrastructure planning. Real estate agents, developers, and financial institutions can utilize either model depending on their forecasting horizon and business goals. Despite the promising findings, this study also acknowledges certain limitations. The models were limited to time-series components without incorporating external variables such as economic indicators, demographic data, or geographical attributes. Incorporating these features through exogenous variable modeling or hybrid deep learning frameworks may improve performance and interpretability. Future research could focus on developing multi-input hybrid models, such as SARIMA-LSTM or ARIMA-XGBoost, which have shown success in other domains.

Additionally, validation using longer historical data and data from multiple Saudi cities could uncover region-specific trends and further generalize the findings. In conclusion, this study demonstrates that the ARIMA model provides more accurate short-term predictions for Saudi real estate prices, while the modified SARIMA model offers better model fit due to its capacity to capture seasonal behavior. The findings reinforce the idea that no single model universally outperforms others, and selection must be aligned with the forecasting objective. The study contributes to the growing body of literature focused on post-pandemic real estate modeling in the Gulf region and provides actionable insights for stakeholders engaged in investment planning, housing policy, and real estate analytics. Continued exploration of hybrid and intelligent forecasting systems is recommended to advance the state-of-the-art in real estate price prediction.

Funding Statement

This study did not receive any external funding.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

Acknowledgments

Not applicable.

I. REFERENCES

1. Abinzano, I., Bonilla, H., & Muga, L. **(2023)**. Duty calls: Prediction of failure in reorganization processes. *The Journal of Risk Finance*, 24, 337–353.
2. Adil, I. H. **(2015)**. A modified approach for detection of outliers. *Pakistan Journal of Statistics and Operation Research*, *XI*, 91–102.
3. Alzahrani, S., & Bhunia, A. (**2025**). FinTech Adoption Intention Among Gen Z in Saudi Arabia: Examining the Serial Mediation of User Innovativeness, Perceived Ease of Use, Trust, and Usefulness. *Qubahan Academic Journal*, *5*(3), 559-579.
4. Al-Wadi, S., Al-Rawashdeh, O. M., ALsinglawi, O., & ABU Dalwein, B. A. **(2022)**. Revenue's forecasting of aqaba ports company using wavelet transform and arima models. *PRZESTRZEN*, 22, 143–160.
5. Mohammed, M. A., Lakhan, A., Zebari, D. A., Abdulkareem, K. H., Nedoma, J., Martinek, R., ... & Tiwari, P. (**2023**). Adaptive secure malware efficient machine learning algorithm for healthcare data. *CAAI Transactions on Intelligence Technology*.

6. Alkhatib, K., Almahmood, M., Elayan, O., & Abualigah, L. **(2022)**. Regional analytics and forecasting for most affected stock markets: The case of gcc stock markets during covid-19 pandemic. *International Journal of System Assurance Engineering and Management*, 13, 1298–1308.
7. Alqahtani, A., Bouri, E., & Vo, X. V. **(2020)**. Predictability of gcc stock returns: The role of geopolitical risk and crude oil returns. *Economic Analysis and Policy*, 68, 239–249.
8. Alshammari, T. T., Ismail, M. T., Hamadneh, N. N., Al Wadi, S., & Jaber, J. J. **(2023)**. Forecasting stock volatility using wavelet-based exponential generalized autoregressive conditional heteroscedasticity methods. *Intelligent Automation & Soft Computing*, 35(3), 2589–2601.
9. Wenyang, D., Zhang, Y., & Dzhamankulov, B. (**2024**). The impact of economic growth and foreign investment on the advancement of e-commerce. *Qubahan Academic Journal*, *4*(4), 112-130.
10. Atsalakis, G. S., & Valavanis, K. P. **(2009)**. Surveying stock market forecasting techniques - part ii: Soft computing methods. *Expert Systems with Applications*, 36(3, Part 2), 5932–5941.
11. Aydin, N., & Yurdakul, G. **(2020)**. Assessing countries' performances against covid-19 via wsidea and machine learning algorithms. *Applied Soft Computing*, *97*, 106792.
12. Bastian, M., Heymann, S., & Jacomy, M. **(2009)**. Gephi: An open-source software for exploring and manipulating networks. In *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*.
13. Bello, I. T., Zhai, S., He, Q., Xu, Q., & Ni, M. **(2021)**. Scientometric review of advancements in the development of high-performance cathode for low and intermediate temperature solid oxide fuel cells: Three decades in retrospect. *International Journal of Hydrogen Energy*.
14. Bhagat, S. K., Tiyasha, T., Al-Khafaji, Z., Laux, P., & Ewees, A. A. **(2022)**. Establishment of dynamic evolving neural-fuzzy inference system model for natural air temperature prediction. *Complexity*, 2022, 1047309.
15. Blažun, H., Kokol, P., & Vošner, J. **(2015)**. Research literature production on nursing competences from 1981 till 2012: A bibliometric snapshot. *Nurse Education Today*, 35(5), 673–679.
16. Boyacioglu, M. A., & Avci, D. **(2010)**. An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the istanbul stock exchange. *Expert Systems with Applications*, 37(12), 7908–7912.
17. Dai, Z., & Zhu, H. **(2023)**. Dynamic risk spillover among crude oil, economic policy uncertainty and Chinese financial sectors. *International Review of Economics and Finance, 83*, 421–450.
18. Dai, Z., Zhu, H., & Zhang, X. **(2022)**. Dynamic spillover effects and portfolio strategies between crude oil, gold and Chinese stock markets related to new energy vehicle. *International Review of Financial Analysis, 109*, 105959.
19. Sayed, O. A., & Eledum, H. **(2021)**. The short-run response of Saudi Arabia stock market to the outbreak of COVID-19 pandemic: An event-study methodology. *International Journal of Finance & Economics*, 26(4), 4857–6487
20. Silva, F., Teixeira, B., Teixeira, N., Pinto, T., & Praça, I. **(2016)**. Application of a hybrid neural fuzzy inference system to forecast solar intensity. In *2016 IEEE 27th International Workshop on Database and Expert Systems Applications (DEXA)* (pp. 161–165). IEEE.
21. Abdullah, R. **(2023)**. *Leveraging XGBoost, Random Forest, and Linear Regression for accurate home price forecasting*.
22. Dubey, A. K., Kumar, A., García-Díaz, V., Sharma, A. K., & Kanhaiya, K. **(2021)**. Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustainable Energy Technologies and Assessments, 47*, 101474.
23. Chen, N. **(2023)**. Visual recognition and prediction analysis of China's real estate index and stock trend based on CNN-LSTM algorithm optimized by neural networks. *PLOS ONE, 18*(2), e0282159.
24. Dong, Z., & Zhou, Y. **(2024)**. A novel hybrid model for financial forecasting based on CEEMDAN-SE and ARIMA-CNN-LSTM. *Mathematics, 12*(16), 2434.
25. Ma, X., Zhu, P., Liu, Q., & Wang, Z. (**2025**). A risk prediction model for real estate corporations using high-target semantic BERT and improved GRU. In *ICASSP 2025 - IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 1–5). IEEE.
26. Taylor, S.J. (**1986**) Modelling Financial Time Series. John Wiley and Sons, Ltd., Chichester.
27. Alshammari, T. T., Ismail, M. T., Hamadneh, N. N., Al Wadi, S., Jaber, J. J., Alshammari, N., & Saleh, M. H. (**2023**). *Forecasting Stock Volatility Using Wavelet-based Exponential Generalized Autoregressive Conditional Heteroscedasticity Methods.* Intelligent Automation & Soft Computing, 35(3), 2589–2601.
28. Machine Learning Valuation in Dual Market Dynamics: A Case Study of the Formal and Informal Real Estate Market in Dar es Salaam. Buildings (**2024**), 14(10), 3172.