International Journal of Advanced and Applied Sciences, 12(12) 2025, Pages: 266-279

Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Prediction of TASI returns using sentiment analysis and hybrid modeling methods: ARIMAX, random forest, and XGBoost

Somaiyah Alalmai *

Finance Department, Faculty of Economics and Administration, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 31 July 2025
Received in revised form
25 November 2025
Accepted 5 December 2025

Keywords:
Financial news sentiment
Saudi stock market
Tadawul index
Machine learning models
Market prediction

ABSTRACT

This study examines the predictive relationship between financial news sentiment and the performance of the Saudi stock market, measured by the Tadawul All Share Index (TASI). A news sentiment index is developed using financial headlines from the Saudi Gazette published between March 2017 and March 2025. The FinBERT model, a natural language processing tool designed for financial text, is used to calculate sentiment scores, which are then averaged on a monthly basis. These sentiment measures are combined with key macroeconomic and market variables, including crude oil prices, interest rates, inflation, exchange rates, and trading volume. For prediction, a hybrid modeling framework is applied, integrating ARIMAX, Random Forest, and XGBoost to capture both linear and nonlinear relationships between TASI returns and sentiment. Model performance is evaluated using root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). The results show that news sentiment and oil price movements have a significant effect on market returns, with important implications for investors, analysts, and policymakers in sentiment-sensitive emerging markets such as Saudi Arabia.

1. Introduction

The stock market is very volatile and may be affected by many factors. Investors themselves belong to the number of the most important factors, as they play the key role in pushing the market in this or that direction through their emotions, decisions, and behavior. The impacts of emotions, personal biases, and individual behavior, as highlighted by Blankespoor et al. (2020), can largely affect the investment decision-making processes made by people, and this may further affect the results in markets. Sentiment has been a key factor in the stock market in the past few years, wherein the emotions of the investor have become a driving force behind any change in the stock market. Although traditional views of financial modeling have focused mainly on fundamental and macroeconomic drivers of stock prices, more recent contributions tend to stress how a large fraction of returns and variability can be attributable to changes in investor sentiment (Nguyen et al., 2025). A decision made by an investor may include both rational thinking and emotions, states Barberis et al. (1998), basically because investors make decisions that rely more on rational thoughts and feelings as opposed to making decisions simply on numbers. This fusion of reason and emotion is what makes the markets uncertain (Barberis et al., 1998). Sentiment moving its asset prices away from fundamental pricing may cause inappropriate capital allocation, poor portfolio construction, and a change in the cost of capital (Smales, 2017).

De Long et al. (1990), who developed the concept of noise trader risk, had long formalized this idea (sentiment can disturb asset prices) and demonstrated that the "unpredictable beliefs of noise traders" can cause asset prices to deviate from their fundamental value even if fundamental risk does not exist. While Keynes (1937) offered early insights into the impact of investor sentiment on financial markets, market practitioners had long acknowledged its power. Lefèvre (2018) acknowledged the importance of investors' emotions and developed tools to gauge extreme optimism and pessimism. These insights laid the foundation for measuring sentiment through proxies. According to Zhou (2018), investor sentiment can be measured in three ways, depending on the data used. The first approach relies on market-based data, such as prices

* Corresponding Author.
Email Address: salalmaee@kau.edu.sa
https://doi.org/10.21833/ijaas.2025.12.024
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-6623-3171

and trading volume. The second approach is based on surveys where investors are directly asked about their expectations. The third approach involves analyzing financial news, social media posts, and other text-based sources to infer investor opinions.

One of the most recognized proxies is by Baker and Wurgler (2006), who developed a comprehensive sentiment index by combining several proxies, namely, "closed-end fund discount, NYSE share turnover, Number of IPOs, first day returns on IPO's, Equity share in new issues, and Dividend premium." The index subsequently became one of the most widely adopted measures for quantifying market sentiment. Their index has proven effective in explaining cross-sectional differences in stock returns, especially for stocks that are difficult to value and to arbitrage, e.g., "small-cap stocks, young firms, highly volatile stocks, non-dividend paying stocks, unprofitable stocks, growth and stocks." However, these sentiment measures often exhibit limited predictive power for aggregate stock market movements. Thus, Huang et al. (2015) address this by modifying the index components and aligning them with aggregate market expectations. Huang et al. (2015) proposed an index that eliminates the common noise component, demonstrating enhanced predictive power in both in-sample and out-of-sample contexts. The index outperforms both traditional sentiment indices and established macroeconomic variables in forecasting stock returns. Their version reveals a negative relationship between sentiment and future market returns. The study underscores the importance of investors' biased beliefs about future cash flow, describing it as the force behind predictive power.

As the connection between investor sentiment and market movements becomes increasingly apparent, exploring this relationship transcends a mere innovative approach; it represents a vital step toward a comprehensive understanding of financial market dynamics. Despite the fact that this relationship has been extensively studied in developed and emerging markets around the world (Barberis et al., 1998; Tetlock, 2007; Kim and Lee, 2022; He et al., 2022; Nguyen et al., 2025; Kräussl and Mirgorodskaya, 2017; Schmeling, 2009), research on media sentiment and its impact on stock markets among the GCC countries remains limited. In particular, the emerging Saudi Arabian stock market (Tadawul) presents an unexplored setting for sentiment-based analysis.

Over the past decade, the Saudi Stock market has undergone substantial reforms, including increased foreign access, privatization initiatives, and regulatory updates under Vision 2030. These reforms have increased the influence of the media on investor expectations and market behavior. Yet, the literature on the relationship between sentiment and stock market performance in the Saudi Arabian context remains scarce, particularly when it comes to high-frequency, text-based measures of sentiment. To current knowledge, no prior research has comprehensively explored the relationship between financial news sentiment and the performance of the Saudi stock market, specifically the Tadawul All Share Index (TASI).

This research addresses this gap. Building on the works of Baker and Wurgler (2006), Huang et al. (2015), Smales (2017), and Barberis et al. (1998), the present research paper us-es recent advances in natural language processing (NLP) to construct a high frequency sentiment index based on financial news headlines, offering a novel and potentially more accurate approach to capturing investor sentiment in the Saudi market. The present research constructs a sentiment index based on financial news headlines, obtained using, Saudi Gazette news outlet, and starting from March 2017 to March 2025. English headlines are processed using the FinBERT sentiment analyzer to extract and compute sentiment scores for financial, economic, and political news. The sentiment scores will be calculated and averaged at a monthly frequency and combined with the respective data in the financial market, including the TASI returns, the trading volume, the oil prices, the exchange rate, the inflation rate, and the interest rate. Also, lagged variables are added to the dataset to obtain the delayed impact of the sentiment on the market behavior.

To perform predictive analysis, the study will combine three complementary models, namely, the ARIMAX model, which uses sentiment as an exogenous regressor in the time-series framework to determine its impact on monthly returns. In addition, Machine learning models such as Random Forest and XGBoost are applied due to their capacity to deal with nonlinear interactions and large data volumes. The models are compared based on both in-sample and out-of-sample statistics. Common regression measures of performance will be applied to model performance, including RMSE, MAE, R^2, and diagnostic tests with normality, autocorrelation, and heteroscedasticity.

This study provides several new contributions to behavioral finance, especially in emerging markets such as Saudi Arabia. It is to study the effects of investor sentiment in the Saudi stock market (TASI) by utilizing high-frequency, news-based sentiment data. This fills one of the most important gaps in the literature, as existing literature has largely studied developed markets that have used survey-based and low-frequency sentiment measures.

A key innovation of this study is the development of a sentiment index utilizing a transformer-based language model, FinBERT, specifically fine-tuned for financial text. The study uses FinBERT applied to English language news headlines to capture real-time investor sentiment with enhanced accuracy and contextual awareness, thereby offering an advantage over traditional lexicon-based approaches. Furthermore, the study introduces a new multi-model prediction framework, which, in the first place, reconciles alternative econometric and machine learning methods. Along with this, the study also presents a multi-model prediction model that integrates conventional econometric and deep learning methods. The ARIMAX model is applied to capture linear dependence and autocorrelation in return and capture sentiment as an exogenous variable. In addition to that, Random Forest and XGBoost models are utilized to detect nonlinear connections between sentiment and market behavior. Lagged variables help test delayed sentiment effects, as part of behavioral finance claims that investors do not react immediately. In general, the study contributes to the literature by showing that such a combination of natural language processing and predictive financial modeling is effective in the context of an emerging market. It also offers a replicable framework of how high-frequency sentiment data can be used in the forecasting of returns and new insights into how media sentiment can be used to determine the behavior of investors in unexplored markets such as Saudi Arabia.

2. Literature review

Fama's Efficient Market Hypothesis (EMH), a cornerstone of traditional finance, suggests that the stock market is rational and provides correct prices, meaning that the current prices of securities are close to their fundamental values. However, anomalies have been observed, and these anomalies have been regarded as the beginning of behavioral finance. This approach challenges EMH by incorporating cognitive and emotional factors. Behavioral finance suggests that markets are not always efficient, accepting people as normal and irrational, and it claims that investors tend to have some psychological and emotional biases that lead to irrationality. The approach recognizes that investors have emotional behaviors and emotional reactions to information, and psychological factors, such as fear, optimism, and uncertainty, can significantly influence asset prices, particularly in sentiment-sensitive markets, thus leading to a distortion of actual values (Shiller, 2003). In today's world, investors are exposed to an environment filled with information, from breaking news and expert analysis to numerous opinions circulating in the media. This constant stream can shape and alter investors' market perceptions and influence their decisions (Chen et al., 2014).

2.1. The role of fintech in financial market analysis

The accelerated development in the FinTech field, especially of such technologies as artificial intelligence (AI), machine learning (ML), and big data analytics, has significantly altered the analysis of financial markets. Such technological advances allow handling of huge volumes of unstructured data, including financial news and social media, to extract information about market sentiment to forecast movements of assets (Griffin et al., 2003). AI and ML tools inherently provide a very powerful ability to discern complex non-linear patterns that econometric models could otherwise fail to capture, providing a richer picture of market dynamics (Fischer and Krauss, 2018; Patel et al., 2015). The sentiment analysis of these FinTech tools directly undermines the strict assumptions of EMH since it presents empirical evidence that investor emotions, as captured by complex algorithms, can in fact push the asset prices out of their fundamental values (Bollen et al., 2011). Such a combination of technology and behavioral finance has created new opportunities for research, as new sentiment indices with higher frequency could be built and more effective predictive models could be created that take into consideration the human irrationality in financial decision-making.

2.2. International and developed market evidence

Although numerous studies have explored the impact of investor sentiment on stock market performance across various global markets, research remains limited in the context of the Gulf region, particularly Saudi Arabia. Schmeling (2009) investigated the impact of investor sentiment on expected stock returns across 18 industrialized countries (Schmeling, 2009). The authors use consumer confidence as a proxy for individual investor sentiment. Building on previous research that focuses on the U.S (see, for instance, Ang et al. (2009) and Griffin et al. (2003)). The authors analyze international data to test the sentiment–return nexus; pooling data from multiple countries enhances the statistical power of the analysis, leading to more reliable estimates (Ang and Bekaert, 2007).

The empirical finding reveals a negative relationship, where higher investor sentiment significantly predicts lower future aggregate stock returns across countries and vice versa. This negative relationship between sentiment and returns also extends to different stock types, such as value, growth, and small-cap stocks, and persists over multiple forecasting horizons. Furthermore, by using cross-sectional regressions, the study offers initial evidence that the strength of the sentiment–return relationship varies by country. The authors argue that the relationship is more pronounced in countries with lower levels of market integrity and in those with cultural characteristics associated with herd behavior and overreaction, as suggested by Chui et al. (2010). These findings underscore the influence of behavioral factors on financial markets.

2.3. Developed markets

Building on this international evidence, several studies have focused specifically on developed markets to examine the role of sentiment more deeply. Kräussl and Mirgorodskaya (2017) examined how media pessimism influences long-term stock market returns and volatility, and if the media sentiment reflects and shapes investor sentiment.

The study constructs a monthly media pessimism index based on the frequency of negative versus positive news in headlines of newspaper articles. Drawing on the underreaction and overreaction hypotheses by Barberis et al. (1998), the authors argue that the effects of negative media coverage on market performance unfold gradually over several months. The results highlight that media pessimism has predictive power, pointing out that when media pessimism increases, it tends to be followed by lower market returns 14 to 17 months later and higher returns 24 to 25 months later. The Granger Causality tests confirm the impact of media pessimism in forecasting future market performance; the results are statistically significant. A final note, the study's media sentiment index adds predictive value beyond used sentiment measures like the "Baker and Wurgler index and the Chicago Board Options Exchange Market Volatility Index (VIX)" (Baker and Wurgler, 2006). Similarly, Tetlock (2007) explored how media sentiment interacts with the stock market by analyzing daily articles from *The Wall Street Journal* column. The empirical evidence reveals that when media coverage is pessimistic, stock prices tend to fall temporarily before adjusting back to their fundamental values. The study highlights that both high and low levels of pessimism are linked to spikes in trading activity. Smales (2017) evaluated several proxies to measure the impact of investor sentiment on asset prices, incorporating data over the period 1990 to 2015. The empirical results are in line with behavioral finance theories, identifying a strong and consistent relationship between investor sentiment and stock returns. Among several proxies, the Implied Volatility Index (VIX) emerges as the most preferred in improving model fit and explanatory power. The causality tests in this research paper reveal that investor fear, captured by the VIX, significantly affects returns across different firm sizes, valuation groups, and industry sectors. Moreover, the study finds that the impact of sentiment is more pronounced during recessions, when investor sentiment is typically at its lowest. The impact is stronger, particularly for stocks that are more susceptible to speculative demand.

A more recent study by Garcia (2025) examined how differences in sentiment expressed in social media platforms (like Twitter/X) and traditional news outlets influence the financial health of companies, utilizing a sample of 1,823 U.S. firms over the period between 2015 to 2015. The study introduces "sentiment divergence" as a proxy for financial distress, that is, when the tone in social media differs from that in traditional news media. The study findings show that stable sentiment divergence reduces default risk, meaning that when there is a steady difference in sentiment between social media and what is reported in the news, companies are less likely to default in the next year in particular, a one standard deviation increase in this divergence reduces the one-year default probability by 7 basis points (0.07%). This aligns with theories that say information diversity improves market efficiency, even if contradictory. On the other hand, when the volatility of sentiment divergence increases, the company's default risk goes up; a one standard deviation increase in the volatility of this divergence increases the default risk by 46 basis points (0.46%). Their findings align with noise trading theories, arguing that investors may trade irrationally, driven by emotions or based on inconsistent information. The study also finds that when institutional investors pay more attention to a company, the risk of financial distress can increase significantly. A one standard deviation increase in institutional attention (measured by Bloomberg's News Heat Index) is linked to an 869-basis point (or 8.69%) increase in probability of default. This finding suggests that when institutional investors focus on a company during times of divergent information, they may react in similar ways. This kind of herding behavior can amplify the company's financial distress. Overall, the study highlights the complex relationship between information, investor sentiment, and corporate risk, offering insights for investors and regulators operating in today's media-driven environment.

2.4. Emerging market dynamics and hybrid modeling approaches

While much of the earlier work has focused on developed economies, recent studies have extended the sentiment-return literature to emerging markets, offering valuable insights into different market dynamics. He et al. (2022) investigated the impact of media sentiment on stock returns in China, particularly on publicly A-listed firms. The results highlight the role media plays in shaping investor sentiment and influencing stock price movements over time, a finding consistent with earlier research on media-based sentiment indices. The study builds a sentiment index based on articles from China's leading financial newspapers. They utilize advanced text analysis tools, including both Word2Vec and dictionary methods, to capture the degree of media coverage positivity or negativity, echoing methods found effective in earlier text-mining and financial prediction research

Further expanding the scope, Nguyen et al. (2025) investigated investor sentiment's influence on stock market returns in Vietnam (VNIndex). The study analyses both short- and long-term effects and builds a sentiment index from over 770,000 Facebook posts (2013–2023). To capture long-run dynamics, the study employs a Vector Error Correction Model (VECM), which reveals that sentiment and stock prices adjust toward a long-term equilibrium after being disturbed by short-term shocks consistent with prior econometric modeling using sentiment indices. Additionally, the study utilizes machine learning techniques, including Decision Trees, SVMs, Neural Networks, Random Forests, Gradient Boosting, and Deep Learning to assess the predictive power of sentiment data. These models perform much better than traditional econometric methods in pre-predicting abnormal returns, which is also corroborated by recent findings within the sentiment-driven forecasting modus operandi (Patel et al., 2015; Zhang, 2003).

Kim and Lee (2022) examined the role of investor sentiment on the Korean stock market, the KOSPI and KOSDAQ returns, in a South Korean context. Basing their data on the adjusted turnover, buy-sell imbalances, and the relative strength index techniques, which overlap the technical and behavioral indicators observed in sentiment literature, the research creates a daily sentiment index. This fact is supported by empirical evidence, and investors' sentiment has a greater influence on the KOSDAQ, which is primarily traded by retail traders. Moreover, the paper examines the firm-level factors underlying the sentiment-return effect, consistent with the earlier findings in the literature on the cross-sectional effects of sentiment (Baker and Wurgler, 2006; Huang and Zhou, 2017). Finally, mobile trading is revealed to benefit emotional investors by enabling them to make more rational choices, an aspect also mentioned in works focused on the effects of digital technologies in shaping behavioral biases and noise trading. These studies provide strong arguments about the effect of the mood of investors in the stock markets, either based on consumer confidence or the tone of media written on social media. These results justify the increasing use of behavioral finance and machine learning methods in the measurement of how sentiment influences the pricing of assets in emerging and advanced economies.

2.5. Evolution and methodologies of sentiment analysis

The investor sentiment is another important element of behavioral finance, which has been captured by different proxies over the years. Earlier methods were based on market data such as the volume of trade and prices, or a direct survey of the investor to find out his or her expectations (Zhou, 2018). The most notable of such is the comprehensive sentiment index of Baker and Wurgler (2006), which was a collection of multiple proxies, including closed-end fund discounts and IPO activity. Though powerful, those measures were sometimes weak predictors of aggregate stock market activity. This index was later refined by Huang et al. (2015) to increase its predictive strength by removing the prevalent noise factors showing an inverse association between sentiment and subsequent market returns.

Lately, there has been a development of interest in text-based sentiment analysis, using the huge amount of information present in the form of financial news, social networking updates, and other textual data. Early text-based techniques tended to be lexicon-based, in which texts were scored against pre-prepared lists of positive and negative words (Hutto and Gilbert, 2014). Such models as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks provided better performance in recognizing sequential data and more intricate linguistic patterns (Fischer and Krauss, 2018; Sharma et al., 2021).

The significant innovation in sentiment analysis with text in the field of finance has been the introduction of transformer-based language models, including BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019). Pre-trained on very large text corpora, these models are good at context and semantic relationships. FinBERT, a BERT variant that is explicitly fine-tuned on financial text, is an important step in that direction (Araci, 2019). The increased accuracy of FinBERT in dealing with financial jargon and contextual nuances ensures that it would outperform both the traditional lexicon-based and general-purpose machine learning methods of capturing real-time investor sentiment (Pankratz, 2012). Nevertheless, despite the use of the more sophisticated models, such as FinBERT, there are still challenges. These limitations are essential to know when it comes to strong sentiment analysis in the financial markets.

2.6. Research gap and current study's contribution

Although much has been documented about the topic of investor sentiment and its effect on the stock markets across the world, there still exists a considerable gap in high-frequency and text-based sentiment analysis in the Saudi Arabian setting. Although substantial reforms in the Saudi stock market (Tadawul) have occurred as part of Vision 2030, which have enhanced the role of media in manipulating investor expectations, the current literature on the connection between sentiment and the performance of stock markets in the specified emerging market is scarce, particularly in the context of high-frequency sentiment measures and text-based sentiment measures. The Saudi market has unique characteristics such as investors, regulatory regime and price sensitivity to crude oil, which offer an unexplored environment of sentiment analysis that cannot be generalized to other emerging markets, such as China, Vietnam or South Korea due to differences in media or investor behavior (He et al., 2022; Kim and Lee, 2022; Nguyen et al., 2025).

As far as it is known today, no other studies have been conducted that would focus on the correlation between the sentiment of the financial news and the behavior of the Tadawul All Share Index (TASI) with the help of sophisticated methods of natural language processing. This paper fills this gap by creating a high-frequency sentiment index using financial news headlines posted by the Saudi Gazette using an NLP tool specifically designed to analyze financial texts, the FinBERT model. In comparison with conventional lexicon-based techniques, this innovative approach includes better accuracy and contextual interpretation that allows for measuring the real-time investor sentiment in the Saudi market more precisely.

Besides, the study presents a multi-model predictive structure that involves the ARIMAX model to include only linear interrelationships and autocorrelation, along with Random Forest and XGBoost models to detect and utilize non-linear connections between sentiment, macroeconomic factors, and market trends. The availability of lagged variables can also enable the analysis of the delayed effect of the sentiment, which can be in accordance with the behavioral finance theories that contend that investors do not always react immediately.

Table 1 presents a summary of the main contributions found in the existing literature.

Table 1: Summary of key literature contributions and identified research gap

Category	Representative studies/proxies	Key findings/methods
General sentiment theory	Keynes (1937), Lefèvre (2018), Barberis et al. (1998), De Long et al. (1990)	Investor emotions influence markets, deviations from fundamentals.
Sentiment measurement proxies	Baker and Wurgler (2006) (closed-end fund discount, IPOs, etc.); Huang et al. (2015) (modified index); Zhou (2018) (market-based, surveys, text-based)	Comprehensive index, improved predictive power; various data sources.
Developed markets (media sentiment)	Kräussl and Mirgorodskaya (2017) (media pessimism, long-term); Tetlock (2007) (Wall Street Journal, temporary price fall); Smales (2017) (VIX, recessions); Garcia (2025) (social media vs. news divergence)	Media influences returns/volatility; sentiment impacts different stock types; divergence as a distress proxy.
Emerging markets (sentiment)	He et al. (2022) (China, media sentiment, short/long-term); Nguyen et al. (2025) (Vietnam, Facebook posts, ML); Kim and Lee (2022) (S. Korea, KOSPI/KOSDAQ, retail traders)	Sentiment impacts return, machine learning effectiveness.
Current study's contribution	Predicting TASI returns based on sentiment analysis and hybrid modeling methods	High-frequency news-based sentiment (FinBERT), hybrid ARIMAX, RF, XGBoost for Saudi TASI.

3. Data and research methodology

3.1. Data sample

To investigate the influence of news sentiment on the performance of the Saudi stock market (Tadawul All Share Index (TASI)), the dataset integrates information about the sentiment scores with macroeconomic indicators. The sentiment index is constructed based on data derived from English-language headlines published by the Saudi Gazette news outlet. The sample incorporates information covering the period from March 2017 to March 2025, a timeframe that includes several market shocks, including both major economic events and routine market fluctuations. The macroeconomic indicators, including Crude Oil prices, Interest Rate, and Inflation Rate data, are fetched via Trading Economics API, and real effective exchange rates (REER) are obtained from the International Monetary Fund (IMF) database. Information for TASI trading volume obtained from investing.com. Also, lagged variables are added to the dataset to obtain the delayed impact of the sentiment on the market behavior. Whereas the dependent variable (Return) is calculated using the daily prices of TASI.

3.2. Methodology

The study uses three prediction models to explore the relationship between the news sentiments and TASI returns. Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) is an Econometric model, and it is used in consideration of both autoregressive dynamics and the effect of exogenous variables like sentiment and macroeconomic indices. This paper employs the use of FinBERT sentiment analyzer, which is used to extract sentiment and compute the sentiment score of the financial, economic, and political news. FinBERT is an artificial intelligence algorithm that carries out sentiment analysis in the financial domain based on the BERT architecture (Devlin et al., 2019). It is a popular natural language processing (NLP) model, which is financial text-based (Jiang and Zeng, 2023). With the help of the Hugging Face Transformers library, the pre-trained model is downloaded, and each headline in the data is assigned a set of sentiment scores (positive, negative, and neutral) (Araci, 2019).

To ensure that it is well aligned with the reporting frequency of macroeconomic indicators, they are summarized at a monthly frequency. It is a systematic conversion to a two-step aggregation process of daily news sentiment scores into a monthly sentiment index. It transforms the sentiment labels (positive, neutral, negative) into weights (1, 0, -1) and multiplies the weights by the respective FinBERT confidence scores to give weighted sentiment scores of both news titles and descriptions. These weighted values are then summed up to a single dataset. The first step in aggregation is to compute the mean daily sentiment by taking the mean of all sentiment scores of several news articles on the same day. The second step involves the average monthly computation of sentiment by summing up these daily averages. The sentiment index is then combined with the respective data in the financial market, including the TASI returns, the trading volume, the oil prices, the exchange rate, the inflation rate, and the interest rate.

Rolling features are also used in this study to capture the short-term market dynamics. The rolling features are obtained by using the sliding 3-month window on returns: The rolling mean is the short-term trend that is obtained by averaging the returns over the three periods, and the rolling standard deviation is the short-term volatility that is obtained by averaging the returns over the three periods. Further, momentum is calculated as the difference between the first and second lag of returns, which represent recent changes in the direction.

The study relies on Sharma et al. (2021), who used ARIMAX to predict stock prices, and Patel et al. (2015), who applied Random Forest to predict whether a stock was going up down in Indian markets. The first model is the ARIMAX model, which estimates the autocorrelation and trends and incorporates the sentiment scores as exogenous variables to assess their impact on the TASI returns. The second model is the Random Forest Regression model, which also accounts for nonlinear relations between sentiment and market indicators. Finally, XGBoost is employed as the third model, and it possesses more predictive power compared to the gradient-boosted decision trees that can handle feature interactions.

3.2.1. ARIMAX

The study uses the ARIMAX model, a time-series econometric model, to determine the relationship between the sentiments of news, macroeconomic factors, and the performance of the TASI index. ARIMAX is a generalization of the univariate ARIMA model, with suitable exogenous regressors, and enables more wide-ranging simulation of the dynamics, and is an extensively applied method of time series forecasting. This helps the model to not only reflect the autocorrelation and moving average tendency in the target variable (TASI returns) but also reflect the impact of other external variables such as sentiment scores, crude oil prices, inflation rates, and trading volume. The ARIMAX model belongs to the most popular tools in financial econometrics, which enables the combination of time series dynamics with fundamental or behavioral data to enhance the accuracy and interpretability of forecasts in volatile markets. Mathematically, the ARIMAX model is a mix of autoregressive (AR), differencing (I), and moving average (MA) models, and the linear combination of external regressors, and may be written as:

$$Y_t = c + \sum_{i=1}^{p} \emptyset_i \, Y_{t-i} + \sum_{j=1}^{q} \theta_j \, \epsilon_{t-j} + \sum_{k=1}^{r} \beta_k \, X_{t,k} + \epsilon_t \quad (1)$$

where,
Y_t = Dependent variable at time 't.'
$\emptyset_i$ = The autoregressive parameters.
θ_j = The moving average part parameters.
$X_{t,k}$ = Exogenous variables at time 't.'
β_k = Coefficients associated with these exogenous variables.

3.2.2. Random forest model

A progressive modeling approach is used to fully simulate the behavior of financial returns. However, given the possibility of nonlinear patterns, interaction effects, and structural changes that ARIMAX may not capture (Zhang, 2003), the study is extended by applying a Random Forest model, which may mimic non-linearities and high-dimensional feature spaces without requiring tight parametric assumptions (Breiman, 2001). Breiman (2001) proposed Random Forest, an aggregate learning approach that uses bootstrap aggregation (bagging) and selection of random features to provide strong predictions for both regression and classification problems. Each tree in the model is trained on a distinct bootstrapped subset of the data and evaluates splits using a random selection of features to reduce variance and overfitting. The final prediction is achieved by averaging the outputs of all trees; this aggregate improves both stability and generalization performance. Genuer et al. (2008) refined the Random Forest algorithm to handle high-dimensional datasets, providing solid feature ranking and variable selection techniques for predictive modeling. Random Forests have been found to be statistically consistent, which shows that as the sample size grows, the model's prediction approaches the true underlying function. It assists in attaining the resilience and generalizability of the model in high-dimensional settings (Scornet et al., 2015). The general formula for the Random Forest model is as follows:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} f_t(x) \quad (2)$$

where,
$\hat{y}$ = Predicted output.
T = The number of trees in total.
$f_t(x)$ = The t^{th} Decision tree prediction.
X = Feature vector of input.

3.2.3. Extreme gradient boosting model (XGBoost)

XGBoost is used in the study after Random Forest to encompass more nonlinear relationships to improve the accuracy of the predictions. Chen and Guestrin (2016) proposed XGBoost (Extreme Gradient Boosting), a novel machine learning algorithm that extends the gradient boosting algorithm with significantly faster regularization and model accuracy. XGBoost is particularly effective on structured data, and is particularly successful in scalable and efficient computing, and has found much utility in financial modeling and forecasting. Compared to classic decision trees or random forests, XGBoost minimizes prediction errors sequentially via gradient descent optimization, and includes regularization (L1 and L2) to avoid overfitting, with the result that it is especially useful with complex, noisy data like financial time series. XGBoost employs a boosting algorithm that sequentially constructs trees where each tree attempts to address the mistakes of the prior tree. This gradient-based optimization captures non-linear, complicated connections in the data.

Furthermore, XGBoost has a feature of missing data handling and a regularization strategy, which is perfect for optimizing model performance. The general equation of XGBoost is represented as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \ f_k \in F \quad (3)$$

where,
$\hat{y}_i$ = Predicted output for instance i.
K = Total number of trees.
$f_k(x)$ = k^{th} regression tree (function).
F = Space of regression trees.
x_i = Input feature vector.

4. Results and discussion

This section covers the findings of the various statistical and machine learning models that were used to forecast the returns of TASI (Tadawul All Share Index) and the interpretation of the same, including the macroeconomic indicators and sentiment analysis. The descriptive statistics and the initial data treatments are used to understand the distribution's features and outlying values. This is followed by the unit root tests to check the suitability of stationarity in each of the variables used to ensure that the time series models are not subject to estimation failures. The gist of the analysis is put into practice and compares the results of three models: ARIMAX, Random Forest, and XGBoost. The comparison of models is done on in-sample and out-of-sample statistics such as RMSE, MAE, and R-squared, and normality, autocorrelation, and heteroscedasticity diagnostics. Through the comparative discussion, the strengths and the limitations of each of the modeling approaches are brought out, and this gives an idea of which of the sentiment and economic variables are more informative indicators of Saudi stock market returns.

4.1. Descriptive statistics

First, descriptive statistics of all variables such as Crude Oil, Real Effective Exchange Rate (REER), Interest Rate, Inflation Rate, Volume (% change), Return, and Weighted Sentiment are estimated to analyze their central tendency, the nature of shapes of the data and dispersion, and reveal possible outliers. The descriptive analysis indicates that most of the variables, such as interest rates, REER, inflation, and crude oil prices, have a moderate variation with nearly normal or flat distributions. Most importantly, it is observed that leptokurtosis (20.46) and a high positive skewness (3.31) are present in Weighted Sentiment, implying the existence of extreme values. In addition, the volume (% change) variable is skewed on the right (1.06), and a slight negative slope in the TASI returns indicates a tendency towards negative returns, stressing that the outlier's treatment is required. On the whole, the majority of the variables included in the dataset demonstrate non-normal characteristics, which means that outlier treatment and distributional adjustments must be done prior to the implementation of the forecasting models.

After identifying outliers using a boxplot graph, variables to be used (including the Crude Oil, Volume (Percent Change), Return, and Weighted Sentiment, among others) are winsorized at the fifth percentile levels to reduce the effects of outliers. The results of the descriptive statistics after the application of winsorization indicate that the effects of extreme values have greatly reduced on the selected variables. The decrease in the standard deviations of the return, crude oil, volume (percentage change), and weighted sentiment reflects less dispersion and increased stability in distribution. The highest cost of crude oil decreased to 96.75 in place of 111.91, as shown in Table 2, and the value is lower at 73.09 compared to 131.84. Similarly, the weighted Sentiment score that in the previous case displayed severe outliers now lies in the range -0.73 to -0.19 with vastly reduced standard deviation (0.15). Further, having made the necessary adjustments, it appears that winsorization can help reduce the effects of outliers and generate more sound and robust data that is further modeled and analyzed.

4.2. Augmented Dickey-Fuller test

The research has used the Augmented Dickey-Fuller (ADF) unit root test to determine the stationarity of all the dependent and independent variables. This was knowing their order of integration and to be sure that their mean, variance, and autocovariance would not change with time. To develop solid models and forecasts relative to volatility, a time series should be stationary (Nelson, 1991).

Table 2: Descriptive statistics

Variables	Observstion	Mean	Std	Min	25%	50%	75%	Max
Crude oil	95	66.64	15.31	39.27	54.28	68.50	77.09	96.75
Real effective exchange rate	95	116.08	3.28	110.63	113.27	115.66	118.83	123.76
Inflation rate	95	0.80	3.01	-5.00	-0.95	1.50	2.50	6.20
Interest rate	95	3.04	1.86	1.00	1.13	2.50	5.00	6.00
Volume % change	95	4.86	31.20	-34.99	-19.72	-3.15	22.51	73.09
Return	95	0.01	0.05	-0.08	-0.02	0.01	0.04	0.08
Weighted sentiment	95	-0.49	0.15	-0.73	-0.59	-0.52	-0.38	-0.19

Table 3 provides the results of the ADF unit root test. One can see that the results of the ADF test indicate that the following variables at the level Return, Weighted Sentiment, Volume (% change), and Real Effective Exchange Rate have p-values below 5 percent. This caused the rejection of the null hypothesis, and the conclusion was that these variables are such that they are stationary both at the level and are of zero-order integrated (I (0)). On the other hand, the Crude Oil, Inflation Rate, and Interest Rate at the level have p-values larger than 5%. When that happened, the results did not reject the null hypothesis of the presence of a unit root in the time series, and it was decided that in such variables, there is a unit root at the level. These non-stationary variables were made to follow a stationary process by performing a first-order difference. At the first difference, p-values of these series turned out to be less than the 5% level of significance, and the null hypothesis was rejected. Thus, the Crude oil, Inflation Rate, and Interest Rate variables were I (1) stationary. Therefore, the ADF Unit root test shows that the crude oil, inflation rate, and interest rate time series are integrated of the first order (I(1)).

Table 3: Unit root test results (ADF)

Variables	Level		1st difference	
	t-stats	p-value	t-stats	p-value
Crude oil	-1.997181	0.60	-9.206757	0.00*
Real effective exchange rate	-3.845801	0.01*		
Inflation rate	-1.731614	0.74	-3.502514	0.01*
Interest rate	-2.636438	0.26	-2.981454	0.04*
Volume % change	-6.415435	0.00*		
Return	-5.098053	0.00*		
Weighted sentiment	-7.014998	0.00*		

*: The level of significance at a 5% confidence interval

4.3. ARIMAX

In this section, the use and outcome of a model utilizing ARIMAX to measure the autoregressive nature of TASI (Tadawul All Share Index) returns and the impact of exogenous explanatory variables are stated. First, by pre-diagnostic tests, both Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots displayed no major peaks after lag 0 predicting that ARIMA (p, d, q) would be optimal at order p = 0, d = 0 and q = 0. But, when this initial ARIMAX specification was fitted we found there was autocorrelation in the residuals. To correct this, autoregressive terms have been included in the model and the result of the different ARIMAX orders has been compared. Out of the estimated models, the order of ARIMAX (1, 0, 0) was the best fit giving an overall good explanation of the autoregressive pattern in addition to the impact of the exogenous predictors. This two-step procedure of evaluation is also consistent with the studies of earlier researchers, including Rahman and Hasan (2017), who compared various ARIMA structures and chose the best model about statistical efficiency. A set of exogenous regressors and the dependent variable (Return) was used to estimate the ARIMAX with an autoregressive order of 1 (ARIMAX (1, 0, 0)) model. In-sample fit of the model was high as the Akaike Information Criterion (AIC) was -369.356, whereas the log-likelihood was 210.678. These are good values and show that the model demonstrated a good in-sample fit, indicating strong relative performance and lower AIC value when compared to other order specifications.

Table 4 discusses the fitting outcome of the ARIMAX (1, 0, 0) model, which includes coefficient estimates, standard errors, z-values, and p-values of each variable, a set of in-sample and out-of-sample evaluation statistics together with important types of diagnostic tests. Exogenous variable significance was analyzed based on p-values whereby at the 5% significance level, crude oil, return rolling mean, and first lag of return rolling mean are the only exogenous variables that are statistically significant predictors of returns, indicating that higher crude oil prices and 3 months rolling return positively influence returns. The positive coefficient of Crude Oil (0.003) indicates that the increase in price of Oil has a positive contribution to the returns of TASI, which is expected of an oil-exporting economy such as Saudi Arabia, where the rise in oil revenue has a positive impact on government expenditures and confidence of investors, this encourages essential businesses such as energy, finance, and construction, hence increasing stock market performances. Hammoudeh and Aleisa (2004) identified a strong link between oil prices and GCC markets. Arouri and Rault (2012) found that oil price changes had a positive impact on stock returns.

The well pronounced positive coefficient 1.828 of Return Rolling Mean clearly demonstrates a meaningful momentum effect, that past average returns exert a significant impact on current returns and its first lag (t-1), the coefficient has a large negative value -1.435 (p-value = 0.000), which exhibits an ill effect and may imply a mean reversion or a complex lag effect. Whereas Interest Rate (p-value = 0.095) and Weighted Sentiment (p-value = 0.067) are also found to be significant at a 10% significance level. Table 4 illustrates a positive coefficient of Weighted Sentiment (0.044), indicating that positive sentiment contributes positively as the positive influence of emotion on stock returns suggests that a bullish public mood boosts investor confidence and buying behavior, resulting in price gains. This impact shows how real-time views of economic or political stability influence market movements. Bollen et al. (2011) found that Twitter mood can predict market movements. Similarly, Nassirtoussi et al. (2014) confirmed the predictive power of emotion in financial markets. On the other hand, the negative value of the coefficient Interest Rate (-0.030) proposes that higher interest rates negatively influence TASI returns, since safer alternatives, such as bonds, provide higher returns and also businesses face higher borrowing cost, which limits their potential to grow and profit, this reduces investor confidence, resulting in lower stock market returns (Bernanke and Kuttner, 2005). In terms of the significance of the autoregressive terms, ar.L1 (p-value = 0.065) is marginally significant at a 10 percent significance level and has a coefficient of -0.268, which represents a slight negative persistence of the return variable that barely has any influence on the current returns despite the exogenous variables. All other exogenous variables, such as Real Effective Exchange Rate, Inflation Rate, Volume % Change, Return Roll Standard Deviation, and their lagged values (t-1 and t-2), are not statistically significant at either the 5 or 10 percent level, implying that they have no direct strong predictive abilities on TASI returns in this model.

Table 4: ARIMAX (1, 0, 0) model estimation results

Variables	Coefficient	Standard error	Z-value	P > |z|
Crude oil	0.003	0.001	4.528	0.000*
Real effective exchange rate	0.000	0.003	0.034	0.973
Inflation rate	0.004	0.003	1.345	0.179
Interest rate	-0.030	0.018	-1.668	0.095**
Volume % change	0.000	0.000	0.419	0.675
Weighted sentiment	0.044	0.024	1.835	0.067**
Return roll mean	1.828	0.213	8.572	0.000*
Return roll std	0.216	0.209	1.034	0.301
Crude oil (t-1)	0.000	0.001	-0.547	0.585
Crude oil (t-2)	-0.001	0.001	-1.255	0.210
Real effective exchange rate (t-1)	0.005	0.005	0.955	0.339
Real effective exchange rate (t-2)	-0.005	0.003	-1.612	0.107
Inflation rate (t-1)	-0.003	0.004	-0.824	0.410
Inflation rate (t-2)	0.000	0.004	-0.049	0.961
Interest rate (t-1)	0.015	0.017	0.879	0.379
Interest rate (t-2)	0.013	0.018	0.751	0.452
Volume % change (t-1)	0.000	0.000	0.989	0.323
Volume % change (t-2)	0.000	0.000	-0.758	0.449
Weighted sentiment (t-1)	-0.032	0.023	-1.415	0.157
Weighted sentiment (t-2)	-0.027	0.020	-1.336	0.182
Return roll mean (t-1)	-1.435	0.258	-5.57	0.000*
Return roll mean (t-2)	0.149	0.215	0.694	0.488
Return roll std (t-1)	-0.369	0.270	-1.367	0.172
Return roll std (t-2)	-0.035	0.213	-0.162	0.871
ar.L1	-0.268	0.145	-1.843	0.065**
	In-sample	Out-sample		
RMSE	0.0233	0.0189		
MAE	0.0193	0.0156		
R^2	0.7318	0.6906		
Ljung-Box autocorrelation test	0.33			
Heteroscedasticity	0.25			
Jarque-Bera	0.68			

*, **: Significance at 5% and 10%, respectively; ar.L1: The first-order autoregressive coefficient of TASI returns

The explanatory and predictive performance of the model is also confirmed by its evaluation measures. The in-sample evaluation measure supports the model's explanatory power with a Mean Absolute Error (MAE) of 0.0193, a coefficient of determination (R^2) of 0.7318, and a Root Mean Squared Error (RMSE) of 0.0233. The results show that 73.18% of the variation in TASI returns is explained by the model. Additionally, by using an out-of-sample approach, the model's predictive ability is further evaluated. The evaluation measures demonstrate the model's strong forecasting ability beyond the training period, with a Mean Absolute Error (MAE) of 0.0156 and a Root Mean Squared Error (RMSE) of 0.0189; both MAE and RMSE show low average prediction error. The coefficient of determination (R^2) of the out-of-sample model is 0.6906, which suggests that approximately 69.06% of the variation in TASI returns is successfully explained by the model in unseen data. The Out-sample R^2 (69.06%) is marginally lower than the In-sample R^2 (73.18%), although this drop in R^2 is to be expected since models tend to perform better on the data they were trained on.

To validate the appropriateness and reliability of the fitted ARIMAX model, the residual diagnostic test results given by the model are evaluated to assess the presence of autocorrelation, normality, and heteroscedasticity in the model's residuals. Diagnostic test results, including the Ljung-Box test and Jarque-Bera test, show no significant autocorrelation in the residuals and indicate that the residuals are normally distributed. Further Heteroscedasticity test authenticates that the error terms are homoscedastic in nature. Collectively, these tests confirm that the model is correctly specified and statistically reliable for forecasting. Hence, by adding important exogenous predictors, the model ARIMAX (1, 0, 0) successfully represented the dynamics of TASI returns. Its applicability for return forecasting is confirmed by its robust performance and reliable residual diagnostics.

4.4. Random forest model

The ARIMAX model is a canonical representation of modeling linear relationships between a time series and exogenous macroeconomic and sentiment variables based on standard time-series assumptions. However, such linear models are unable to detect non-linearity, effects of interactions, and structural changes (Zhang, 2003). Therefore, the current extension brings Random Forest, which, based on the use of the randomized decision trees, could represent complex nonlinearities without requiring stringent parametric restrictions (Breiman, 2001). The study evaluated multiple train/test splits to find the best test size for the random forest model. Rolling features (such as rolling standard deviation and mean) and lag returns are utilized to improve forecast accuracy. After adding these temporal characteristics, the Random Forest model's performance significantly improved as RMSE and MAE reduced to 0.0273 and 0.0236, respectively. As shown in Table 5, the model's R^2 also improved to 0.3543, revealing important patterns and accounting for almost 35% of the variance in returns. These findings underscore the relevance of feature engineering, particularly time-series transformations, in enhancing model performance when applying machine learning techniques to financial information.

Furthermore, to validate the reliability of the model, various post-diagnostic tests are conducted. The Shapiro-Wilk test for normality gives a statistic of 0.9635 and a p-value of 0.7804, indicating that the null hypothesis of normality cannot be rejected. This implies that the residuals are roughly normally distributed, which meets one of the critical requirements for model dependability. Further, as in Table 5, the heteroscedasticity test with a p-value of 0.3738, indicating that the null hypothesis of homoscedasticity is valid and cannot be rejected. This shows that the residuals have constant variance. Moreover, for detecting the autocorrelation Durbin Watson test is conducted, the test statistic is 2.1279, which is around the optimal value of 2. This suggests that the residuals are white noise. Overall, the post-diagnostic findings show that the residuals are normally distributed, homoscedastic, and uncorrelated, which validates the model's statistical soundness.

Table 5: Random Forest evaluation metrics

Evaluation metrics	Value	
RMSE	0.0273	
MAE	0.0236	
R^2	0.3543	
	Statistic	p-value
Shapiro-Wilk test	0.9635	0.7804
White test for heteroscedasticity	14	0.3738
Durbin-Watson statistic	2.1279	–

4.5. XGBoost model

In this study, after Random Forest, XGBoost is used to capture more nonlinear relationships and improve prediction accuracy. XGBoost (Extreme Gradient Boosting) is applied to improve speed, regularization, and model accuracy. Results in Table 6 showed that the model performs well by adding the rolling mean and standard deviation feature, showing an increment in the predictive power with an MAE value of 0.0184, RMSE of 0.0231, and R^2 of 0.5382. These findings indicate that the model accounts for about 53.82% of the variance in returns, indicating a good predictive power. According to the results, the most dominant variables are changes in crude oil prices, lagged return values, rolling mean and volatility of returns, sentiment scores, and macroeconomic indices like inflation and volume fluctuations.

Significantly, the difference of crude oil and rolling return mean emerged as the most powerful predictors, highlighting the importance of short-term momentum, market emotion, and macroeconomic variables in forecasting future return orientations. To validate the reliability of the XGBoost model, various post-diagnostic tests are conducted. Firstly, the Shapiro-Wilk test is performed for checking the normality, which gives a statistic of 0.9060 and a p-value of 0.1378, indicating that the null hypothesis of normality cannot be rejected. This implies that the residuals are roughly normally distributed, which meets one of the critical requirements for model dependability. Further, as in Table 6, the heteroscedasticity test with a p-value of 0.3738, indicating that the null hypothesis of homoscedasticity is valid and cannot be rejected. This shows that the residuals have constant variance. Moreover, for detecting the autocorrelation Durbin Watson test is conducted, the test statistic is 2.0809, which is around the optimal value of 2. This suggests that the residuals are white noise.

Table 6: XGBoost evaluation metrics

Evaluation metrics	Value	
RMSE	0.0231	
MAE	0.0184	
R^2	0.5382	
	Statistic	p-value
Shapiro-Wilk test	0.9060	0.1378
White test for heteroscedasticity	14	0.3738
Durbin-Watson statistic	2.0809	–

4.6. Combined model performance evaluation

The findings indicate that ARIMAX provides the best fit (R^2 = 0.73 in-sample, 0.69 out-of-sample); however, such high explanatory power of financial returns is probably due to overfitting since the autoregressive dynamics are used, and numerous insignificant lagged variables are included. In comparison, machine learning models have more realistic estimates that are cautious. XGBoost performs better than Random Forest (R^2 = 0.54 vs. 0.35) because its boosting structure can capture nonlinear interactions better than the averaging process of Random Forest. Notably, the reduced R^2 of these models can provide a more accurate perspective of the inherent noise in financial data, as opposed to reflecting inferiority.

In general, the results indicate that ARIMAX performs well in harnessing autocorrelation and finding important drivers (oil, sentiment, interest rates, momentum), whereas XGBoost is strong in nonlinear predictive capability. Collectively, these models provide complementary information about the determinants of TASI returns and the contribution of investor sentiment to the Saudi market.

Following the model comparison in Table 7, the feature importance analysis (Fig. 1) provides more insights into the drivers of prediction accuracy. The findings indicate that the most influential predictor is crude oil price differences, which highlight the sensitivity of the Saudi market to global energy price changes. Rolling return measures (mean and volatility) and return lags are also among the highly ranked, and they demonstrate the presence of autocorrelation and short-term momentum effects in TASI returns.

Notably, weighted sentiment is a mid-level predictor, which lies near the lags of returns. This implies that investor sentiment is not the most important driver, but it has a significant complement to market fundamentals, in the sense that it provides a measure of psychological dynamics that are typically ignored in conventional econometric models. In comparison, macroeconomic variables like exchange rate, inflation, and interest rates have comparatively low explanatory power when direct return dynamics and crude oil effects are incorporated.

Table 7: Model comparison table

Model	RMSE	MAE	R^2
Random Forest	0.0273	0.0236	0.3543
XGBoost	0.0231	0.0184	0.5382
Arimax (in-sample)	0.0233	0.0193	0.7318
Arimax (out-sample)	0.0189	0.0156	0.6906

Altogether, the findings of the feature importance indicate that the previous findings in model performance indicate that predictive power is concentrated on several strong variables, and sentiment adds extra, but less important value to the analysis of the return variation.

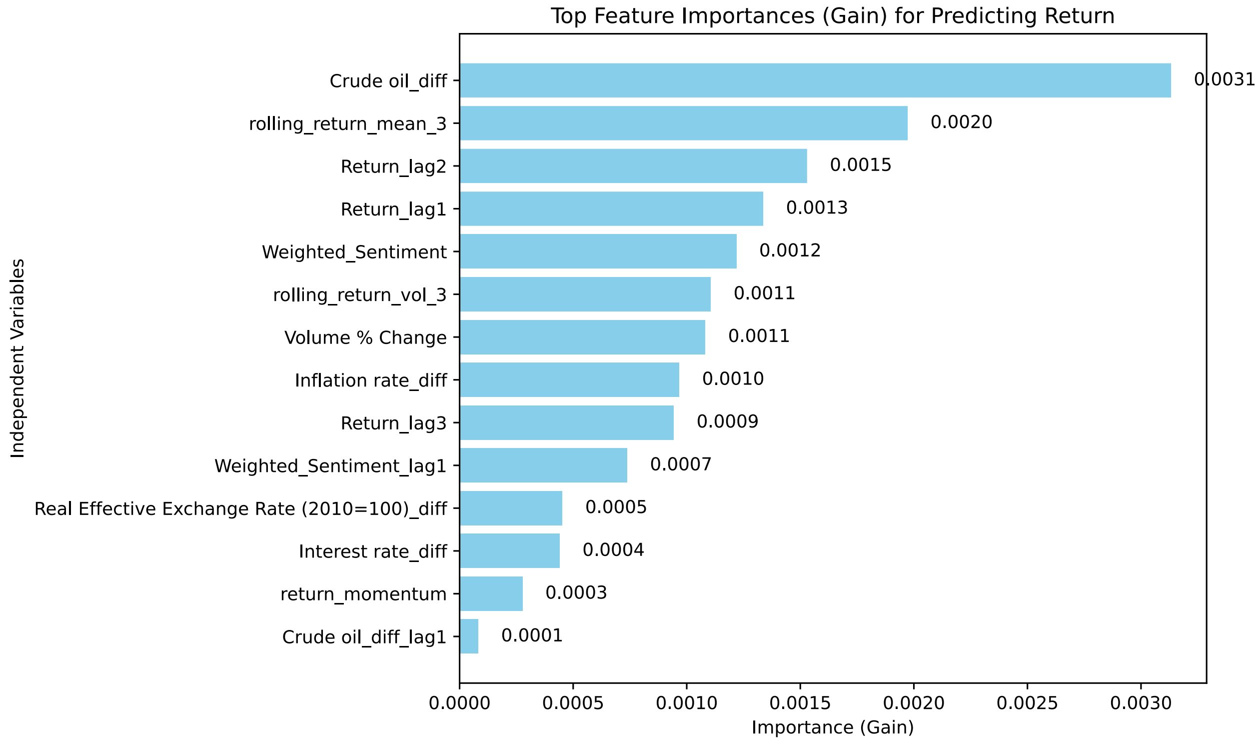

Fig. 1: Feature importance graph

5. Conclusion

This study investigated three models including one time series (ARIMAX) and machine learning models (Random Forest and XGBoost) for investigating the relationship between Saudi stock market performance (returns of TASI) and news sentiments. The study utilized key performance measures including RMSE, MAE, and R^2, as well as diagnostic tests to ensure the reliability of the model. The theory of linear time series is that the ARIMAX model has excellent in-sample accuracy and decent out-of-sample performance. The model can be applied because it can be interpreted, and it is efficient when the relationships are linear and familiar. It is, however, plagued by non-linear effects and higher-order complexities (Pankratz, 2012). The rolling features only had moderate improvements on the Random Forest model (RMSE: 0.0273, R^2: 0.3543). It aligns with the findings of such studies as Lahmiri and Bekiros (2020), which show that Random Forests can be used to model non-linear dependencies but may not work well unless they are tuned or time sensitive. The XGBoost model was better than the Random Forest model, with lower RMSE (0.0231), MAE (0.0184), and higher R^2 (0.5382). These results correlate with the previous literature (Chen and Guestrin, 2016; Fischer and Krauss, 2018) that pointed out the outstanding predictive performance of XGBoost due to its gradient boosting model and powerful regularization strategies, as well as the ability to learn complex nonlinear relationships in financial data. The ARIMAX model has the highest R^2, though this may be an overfitting problem, because it only grasps linear trends and may not generalize to nonlinear dynamics. Although XGBoost has a slightly lower R^2, it is more convenient to reflect complex, real-world financial relationships. As stated by the previous research and practical data, the XGBoost model is the most effective of the three approaches.

The results demonstrated that some external factors have a significant impact on Saudi stock market performance. Crude oil, return rolling mean, and first lag of return rolling mean are the only exogenous variables that are statistically significant predictors of returns, highlighting the importance of revenue from oil and short-term momentum in influencing investor behavior in an oil-dependent economy. Whereas, Interest rate and weighted sentiment are also statistically significant predictors of returns at a 10% significance level, indicating that positive sentiment positively boosts investor confidence and buying behavior, resulting in price gains. While higher interest rates negatively influence TASI returns, since safer alternatives provide higher returns and businesses face higher borrowing costs, which limits their potential to grow and profit, this reduces investor confidence, resulting in lower stock market returns. The marginal importance of the autoregressive lag supports the notion of limited short-term memory in return patterns. These findings highlight the relevance of macroeconomic indicators, investor emotion, and momentum for projecting returns in the Saudi market.

List of abbreviations

ACF	Autocorrelation function
ADF	Augmented Dickey–Fuller test
AI	Artificial intelligence
AIC	Akaike information criterion
AR	Autoregressive
ar.L1	The first-order autoregressive coefficient of TASI returns
ARIMA	Autoregressive integrated moving average
ARIMAX	Autoregressive integrated moving average with exogenous variables
BERT	Bidirectional encoder representations from transformers
EMH	Efficient market hypothesis
FinBERT	Financial bidirectional encoder representations from transformers
GCC	Gulf Cooperation Council
I(0)	Integrated of order zero
I(1)	Integrated of order one
IMF	International Monetary Fund
IPOs	Initial public offerings
KOSDAQ	Korean securities dealers automated quotations
KOSPI	Korea composite stock price index
LSTM	Long short-term memory
MA	Moving average
MAE	Mean absolute error
ML	Machine learning
NLP	Natural language processing
PACF	Partial autocorrelation function
REER	Real effective exchange rate
RF	Random forest
RMSE	Root mean squared error
RNNs	Recurrent neural networks
TASI	Tadawul all share index
VECM	Vector error correction model
VIX	Implied volatility index
XGBoost	Extreme gradient boosting

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Ang A and Bekaert G (2007). Stock return predictability: Is it there? The Review of Financial Studies, 20(3): 651-707. https://doi.org/10.1093/rfs/hhl021

Ang A, Hodrick RJ, Xing Y, and Zhang X (2009). High idiosyncratic volatility and low returns: International and further U.S. evidence. Journal of Financial Economics, 91(1): 1-23. https://doi.org/10.1016/j.jfineco.2007.12.005

Araci D (2019). FinBERT: Financial sentiment analysis with pre-trained language models. Arxiv Preprint Arxiv:1908.10063. https://doi.org/10.48550/arXiv.1908.10063

Arouri MEH and Rault C (2012). Oil prices and stock markets in GCC countries: Empirical evidence from panel analysis. International Journal of Finance & Economics, 17(3): 242-253. https://doi.org/10.1002/ijfe.443

Baker M and Wurgler J (2006). Investor sentiment and the cross-section of stock returns. The Journal of Finance, 61(4): 1645-1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Barberis N, Shleifer A, and Vishny R (1998). A model of investor sentiment. Journal of Financial Economics, 49(3): 307-343. https://doi.org/10.1016/S0304-405X(98)00027-0

Bernanke BS and Kuttner KN (2005). What explains the stock market's reaction to Federal Reserve policy? The Journal of Finance, 60(3): 1221-1257. https://doi.org/10.1111/j.1540-6261.2005.00760.x

Blankespoor E, deHaan E, and Marinovic I (2020). Disclosure processing costs, investors' information choice, and equity market outcomes: A review. Journal of Accounting and Economics, 70(2-3): 101344. https://doi.org/10.1016/j.jacceco.2020.101344

Bollen J, Mao H, and Zeng X (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1): 1-8. https://doi.org/10.1016/j.jocs.2010.12.007

Breiman L (2001). Random forests. Machine Learning, 45: 5-32. https://doi.org/10.1023/A:1010933404324

Chen H, Chong TTL, and She Y (2014). A principal component approach to measuring investor sentiment in China. Quantitative Finance, 14(4): 573-579. https://doi.org/10.1080/14697688.2013.869698

Chen T and Guestrin C (2016). XGBoost: A scalable tree boosting system. In the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, San Francisco, USA: 785-794. https://doi.org/10.1145/2939672.2939785

Chui AC, Titman S, and Wei KJ (2010). Individualism and momentum around the world. The Journal of Finance, 65(1): 361-392. https://doi.org/10.1111/j.1540-6261.2009.01532.x

De Long JB, Shleifer A, Summers LH, and Waldmann RJ (1990). Noise trader risk in financial markets. Journal of Political Economy, 98(4): 703-738. https://doi.org/10.1086/261703

Devlin J, Chang MW, Lee K, and Toutanova K (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, USA, 1: 4171-4186. https://doi.org/10.18653/v1/N19-1423

Fischer T and Krauss C (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2): 654-669. https://doi.org/10.1016/j.ejor.2017.11.054

Garcia J (2025). Beyond the headlines: Sentiment divergence and financial distress. Global Finance Journal, 66: 101126. https://doi.org/10.1016/j.gfj.2025.101126

Genuer R, Poggi JM, and Tuleau C (2008). Random Forests: Some methodological insights. Arxiv Preprint Arxiv:0811.3619. https://doi.org/10.48550/arXiv.0811.3619

Griffin JM, Ji X, and Martin JS (2003). Momentum investing and business cycle risk: Evidence from pole to pole. The Journal of Finance, 58(6): 2515-2547. https://doi.org/10.1046/j.1540-6261.2003.00614.x

Hammoudeh S and Aleisa E (2004). Dynamic relationships among GCC stock markets and NYMEX oil futures. Contemporary Economic Policy, 22(2): 250-269. https://doi.org/10.1093/cep/byh018

He Y, Qu L, Wei R, and Zhao X (2022). Media-based investor sentiment and stock returns: A textual analysis based on newspapers. Applied Economics, 54(7): 774-792. https://doi.org/10.1080/00036846.2021.1966369

Huang D and Zhou G (2017). Upper bounds on return predictability. Journal of Financial and Quantitative Analysis, 52(2): 401-425. https://doi.org/10.1017/S0022109017000096

Huang D, Jiang F, Tu J, and Zhou G (2015). Investor sentiment aligned: A powerful predictor of stock returns. The Review of Financial Studies, 28(3): 791-837. https://doi.org/10.1093/rfs/hhu080

Hutto C and Gilbert E (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1): 216-225. https://doi.org/10.1609/icwsm.v8i1.14550

Jiang T and Zeng A (2023). Financial sentiment analysis using FinBERT with application in predicting stock movement. Arxiv Preprint Arxiv:2306.02136. https://doi.org/10.48550/arXiv.2306.02136

Keynes JM (1937). The general theory of employment. The Quarterly Journal of Economics, 51(2): 209-223. https://doi.org/10.2307/1882087

Kim Y and Lee KY (2022). Impact of investor sentiment on stock returns. Asia-Pacific Journal of Financial Studies, 51(1): 132-162. https://doi.org/10.1111/ajfs.12362

Kräussl R and Mirgorodskaya E (2017). Media, sentiment and market performance in the long run. The European Journal of Finance, 23(11): 1059-1082. https://doi.org/10.1080/1351847X.2016.1226188

Lahmiri S and Bekiros S (2020). Intelligent forecasting with machine learning trading systems in chaotic intraday Bitcoin market. Chaos, Solitons & Fractals, 133: 109641. https://doi.org/10.1016/j.chaos.2020.109641

Lefèvre E (2018). Reminiscences of a stock operator. Courier Dover Publications, Mineola, USA.

Nassirtoussi AK, Aghabozorgi S, Wah TY, and Ngo DCL (2014). Text mining for market prediction: A systematic review. Expert Systems with Applications, 41(16): 7653-7670. https://doi.org/10.1016/j.eswa.2014.06.009

Nelson DB (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59(2): 347-370. https://doi.org/10.2307/2938260

Nguyen HH, Ngo VM, Pham LM, and Van Nguyen P (2025). Investor sentiment and market returns: A multi-horizon analysis. Research in International Business and Finance, 74: 102701. https://doi.org/10.1016/j.ribaf.2024.102701

Pankratz A (2012). Forecasting with dynamic regression models. John Wiley & Sons, Ltd., Hoboken, USA.

Patel J, Shah S, Thakkar P, and Kotecha K (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications, 42(1): 259-268. https://doi.org/10.1016/j.eswa.2014.07.040

Rahman A and Hasan MM (2017). Modeling and forecasting of carbon dioxide emissions in Bangladesh using autoregressive integrated moving average (ARIMA) models. Open Journal of Statistics, 7(4): 560-566. https://doi.org/10.4236/ojs.2017.74038

Schmeling M (2009). Investor sentiment and stock returns: Some international evidence. Journal of Empirical Finance, 16(3): 394-408. https://doi.org/10.1016/j.jempfin.2009.01.002

Scornet E, Biau G, and Vert JP (2015). Consistency of random forests. The Annals of Statistics, 43(4): 1716–1741. https://doi.org/10.1214/15-AOS1321

Sharma A, Tiwari P, Gupta A, and Garg P (2021). Use of LSTM and ARIMAX algorithms to analyze impact of sentiment analysis in stock market prediction. In: Hemanth J, Bestak R, and Chen JIZ (Eds.), Intelligent data communication technologies and internet of things. Lecture notes on data engineering and communications technologies, 57: 377-394. Springer, Singapore, Singapore. https://doi.org/10.1007/978-981-15-9509-7_32

Shiller RJ (2003). From efficient markets theory to behavioral finance. Journal of Economic Perspectives, 17(1): 83-104. https://doi.org/10.1257/089533003321164967

Smales LA (2017). The importance of fear: Investor sentiment and stock market returns. Applied Economics, 49(34): 3395-3421. https://doi.org/10.1080/00036846.2016.1259754

Tetlock PC (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3): 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Zhang GP (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50: 159-175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhou G (2018). Measuring investor sentiment. Annual Review of Financial Economics, 10(1): 239-259. https://doi.org/10.1146/annurev-financial-110217-022725