

An Integrated Technical and Fundamental Stock Analysis Model for the Saudi Stock Market (Tadawul)

1. Abstract

This study creates and implements a **predictive machine learning model** to facilitate the possibility of short-term investment intentions within the Saudi Stock Market (Tadawul). Such a demand is necessary due to the inherent complications of researching various financial data/technical relevance/regulatory compliance simultaneously. Many investments would derive value from more rapid, consolidated findings regarding when to enter and how many companies would present ideal recommendations for investor satisfaction. Thus, the model provides a macro-level outlook to facilitate novice-investor stakeholders without the cumbersome, time-consuming necessity of complicated analyses through rapid, generative predictive analytics that is all-inclusive yet efficient.

Empirical **analytical-experimental data** is found within historical data for training purposes with comprehensive results for functional viability. Information derives from company **fundamental financial statements** and an equalized selection of technical indicators - company open, company close, and different technical oscillators (RSI, MACD, Bollinger Bands) as the selected technical indicators - and the binary predictive outcome determines whether 5% profit with 2% stop-loss is anticipated from day-to-day investment intentions. Furthermore, information collected and assessed pays no attention to sentiment analysis or media speculation for the same reason as with any investment intentions within the Saudi Capital Market - as per Saudi Capital Market Authority (CMA) regulations, unlicensed analysis is expressly prohibited, and the literature review notes why such an assessment is necessary.

Preliminary results find that this model has a highly complex ability to assess investment intentions based on correct findings and that the hybridized approach to applicable fundamental statements and technical indicators is a more suitable basis for predictions than unrelated ventures from external findings. Generalizations from this discovery support the specialization of this study as accurate conclusions can be drawn from legitimate investigation relative to learning about the true Saudi market.

The study ultimately concludes that AI-fueled potential predictions based solely on publicly available and quantifiable information are a significant reliable measurement for small investment intentions and valuable proactive suggestions without the need for interim human analysis/emotional readings from potential news and inappropriate information.

2. Introduction

2.1. Research Background and Problematic

The Saudi Stock Market (Tadawul) is the largest and most dynamic market in the MENA region, receiving growing interest from global institutional investors and recent **Saudi Vision 2030** aims at diversifying the economy even further. Such rapid expansion necessitates the use of advanced quantitative solutions relative to increased volatility at a more complex level. Given that investments are only sometimes what they seem, value-based solutions derived from a mere

presentation of historical data or human deduction fall short of encompassing complex time-series developments with predictive, nonlinear features characteristic of **emerging markets**. Accordingly, Artificial Intelligence (AI) and Machine Learning (ML) solutions are more frequently adopted to create predictive data which can foster feasible resolution.

2.2. Dual Predictive Deficiency

To ensure sound investments, two dimensions must be considered throughout investment: **Value** and **Timing**. **Value** is created through fundamental analyses (P/E ratio, ROE, etc.) which consider what one must buy based on intrinsic features as opposed to learned features from price movements which ultimately defer to **Timing**, created through technical analysis which determines when one should buy/sell to generate the expected return on investment for a short-term investment.

Regrettably, too many quantitative solutions exist to fail to create a comprehensive understanding of Value and Timing

1. **Failure to Analyze All Components:** Many deep learning frameworks attempt to forecast stock prices in the short run based on Time (technical analysis) - without a means to reconcile Value Proposition or even an understanding of why some companies may not even have sustainable management teams for any transaction to make sense. This allows potentially high-risk ventures to occur without a fundamental reason to implement them.
2. **Regulatory Non-Compliance Risk:** Many academic papers attempt to supplement predictive prowess with unstructured data - news sentiment or social media sentiment - which involves regulatory compliance risk, contrary to the enforced **Market Conduct Regulation** issued by the Saudi **Capital Market Authority (CMA)** which expressly prohibits agents from communicating unintended and/or subjective market sentiment.

2.3. Motivation Research Objective and Hypothesis

The significance of the study comes from the need for a reliable, regulated, and comprehensive solution for novice-investor stakeholders²². Currently, investors are left to battle with convoluted assessments that take too much time and become frustrating.

As a result, the research seeks to fill the dual deficiency from either an analytical or regulatory approach by introducing a new, reputable, CMA-compliant AI solution that successfully integrates **Value** and **Timing** to create a feasible buy signal (+5% target -2% stop-loss), suggesting the market demand for such integrative prediction.

To validate the suggested hypothesis, advanced **Stacking Ensemble Architecture** should outperform other non-integrated models and the benchmark through statistically and financially driven results - where LSTM (for Timing) has a longitudinal - temporal - predominance while XGBoost (for Value) has a cross-sectional - analytical - depth - to justify not only successful trade results but no violations in the process.

2.4. Research Paper Structure

The remainder of the paper is organized as follows: Chapter 3 provides a relevant literature review to understand existing gaps in the research relative to proper regulations. Chapter 4 explains the proposed Stacking Ensemble Methodology and Model Architecture relative to Dynamic Implementation. Chapter 5 includes backtesting and performance review sections supported by comparable assessments. Chapter 6 concludes the project with significant findings.

3. Related Work

This section attempts to take the reader on a journey through relevant studies based on stock prediction for emerging markets and, more specifically, Saudi Arabia, so that a relative context can be formed for the hybrid model in question to either fill in research gaps or learn from peer efforts.

3.1. Findings Overview

1. Saudi Arabia Stock Prices Prediction Using Machine Learning (2025)

The undertaking is the simplest venture from which subsequent validation and benchmarking is drawn. The aim of this venture is an effort to build a reliable model for the next day predictions of the **Saudi Stock Market (Tadawul)**. The efforts took four classical Machine Learning (ML) algorithms and produced and compared results: **Linear Regression (LR)**, **Random Forest**, **XGBoost**, and **AdaBoost**. The findings emerged from four companies with the greatest price history all within one company's stock share, as well as across the most traded contributors. Yet the most substantial finding was the confidence that the simplest model of all - **Linear Regression** - was the most effective in its simplicity one company's coefficient of determination (R Squared) was equal to **0.9824**. This means that there's a significant likelihood that absolute predictions could be created with LR to act as a base prediction for quantitative endeavors. The gap in academic literature revolves around the exclusively price-based finite approach without any other intentions or fundamental values applied.

2. Predicting Close Price In Emerging Saudi Stock Exchange: Time Series Models (2022)

The venture aimed to expand on similar research ventures in the Tadawul market to see if DL models might be more appropriate than time series models. This model was the **Long Short-Term Memory (LSTM)** network compared to a traditional **Multilayer Perceptron (MLP)** model based on historical price data of the Tadawul. Findings were definitive that the **LSTM model** is more than appropriate historical findings show that its ability to predict sequential price prediction is higher than expected research from this study determined a **correlation coefficient of greater than 0.9950**. Therefore, this model can be compared as a requisite, baseline and effective time series field-based model. The limitation of this study is that, like the prior paper, it derives its findings from a technical/quantitative approach only without any sentiment or other fundamentals to compare.

3. Saudi Market Stock Price Prediction Using Deep Learning Algorithms and Weighted Twitter Sentiment Analysis (2024)

This unique study is the most recently from 2024 and the most comprehensive undertaking to combine a non-structured data source within prediction efforts for the Tadawul. This was an effort to determine if weighted sentiment analysis could be combined with historical data sentiment analysis to achieve success for stock price prediction. The methods themselves were hybrid but apart: a **LSTM model** generated through a purely numerical approach determined **98% accuracy** (Mean Squared Error) while a Support Vector Machine (SVM) was applied to determine polarity of sentiment with **99% accurate polarity** determination. The main point of innovation was to weight each tweet based on how many retweets, favorites and comments were applied. However, this effort has significant limitations - first, Twitter is an informal source with major regulatory issues in Saudi Arabia - and second, without a true integrated fusion-based mechanism, a combined analysis could not determine how much sentiment improved accuracy.

4. A Hybrid Data Analytics Framework with Sentiment Convergence and Multi-Feature Fusion for Stock Trend Prediction (2022)

This paper is most up to date because it reconciles the fusion of features, a pitfall of previous hybrid approaches.

It was applicable to the Dubai Financial Market (DFM).

The architecture is a hybridized framework, with Convolutional Neural Networks (CNN) a means of sentiment analysis to extract features from official news texts and Bidirectional Long Short-Term Memory (BiLSTM) expected for powerful temporal analysis. The most innovative aspect is Sentiment Convergence and Multi-Feature Fusion to avoid sentiment dispersion, with the most impressive finding performance gains against the benchmarks,

- The Base Code Prediction Accuracy: pre-enhancement benchmark prediction accuracy is 0.646.

- The Proposed Prediction Accuracy: The advanced proposal accuracy is 0.785 (an overall increase of 21.4% and an increase of 25.6% to the highest deviation prediction of 0.812 for integrated fusion power).

- Accuracy of 0.785 is 7.8% to 11.6% greater across the board compared to similar fusion approaches with predictive capability.

The paper's significant non-advantage from an academic perspective is that it only focuses on trend prediction without regard for stock's fundamental nature which adds value to the stock and comes from fundamental analysis.

3.2. Summary of Findings and Research Gaps

The findings thus are a general body of literature which combine a comparably similar yet unresolved pattern across all offered solutions of Tadawul so far as they either predict a **timing recommendation without predicting fundamental value** (purely technical, non-compliant models) or **sacrifice compliance** to a subjective, informal sentiment analysis.

This gap was solidified through the mentioned research gaps from literature review. Dardakeh's (2022)} study is not applicable to the Saudi market due to reliance on non-compliant external market data Al-Nefaiand Aldhayani (2022)} and Khuraishiand Alharbi (2025) put forth powerful models, however are **fundamentally lacking** a value-based analysis of intrinsic value and any conditional decision rule for traders Daori et al. (2024) make statistically accurate claims, however focus on statistical error instead of risk-adjusted returns and international, inappropriate data sources for a Saudi inquiry.

Thus, a major **research need** exists for this dual deficiency.

Our study suggests a **structurally coherent, economically coherent, time-sensitive prediction engine** with a novel Stacking Ensemble framework to blend the value-based assumptions of **XGBoost (for Value)** and temporal strength of **LSTMs (for Timing)** as neither have been proposed with the intention of consistent risk-defined conditional signaling.

3.3. Comparative Methodological Analysis Table

Study	Year	Market	Models	Critical Gap
SAUDI ARABIA STOCK PRICES PREDICTION...	2025	Saudi Market (Tadawul)	Comparison of classical ML algorithms (LR, XGBoost, RF, AdaBoost)	Exclusively Technical Data (Timing) without Fundamental Value considerations
Predicting Close Price in Emerging Saudi Stock Exchange...	2022	Saudi Market (Tadawul)	Comparison of Deep Learning models LSTM vs. MLP (exclusively from historical price data).	Pure Technical Price Analysis only, without any non-financial variables (Sentiment).
Saudi Market Stock Price Predictions... and Weighting Twitter Sentiment Analysis	2024	Saudi Market (Tadawul)	Discretionary Hybrid model LSTM for numerical data + SVM for Weighted Twitter Sentiment).	Regulatory Issue from using unofficial sources no proper structure for fusion.
A Hybrid Data Analytics Framework... for Stock Trend Prediction	2022	Dubai Financial Market (DFM)	Advanced Hybrid CNN ,BiLSTM with Sentiment Convergence and Multi-Feature Fusion.	Inapplicable for Saudi Context (different market) Trend Prediction only, excluding Intrinsic Value.

4. Proposed Methodology

4. Proposed Methodology

The methodology applies to daily data between **2021 to 2025** for a window of a sample of Tier 1 companies from Tadawul.

4.1. Base Learners Justification

Base learners are functionally separated between **LSTM** and **XGBoost** as follows:

- **LSTM:** More beneficial in sequential information which means it can rely on long-term dependencies of the 60 days of previous prices. It also relates to **Timing**, dependent on short-to-medium momentum which means more so than there is buy signal momentum the greater chance to assess it.
- **XGBoost:** More helpful in organized, tabled data (financial ratios). It acknowledges the feature importance and helps attribute why certain variables facilitate more non-linear relationships to assess intrinsic value.

4.2. Data and Features Creation

The data is collected from trusted sites to be cited below (**Yahoo Finance**) and created between **2021-2025**.

A. Fundamental Financial Features - The XGBoost Model: Features that assess **Intrinsic Value** (Value):

- **P/E Ratio (Price-to-Earnings):** Implied (from the Income Statement and from the market).

$$P/E = \frac{\text{Daily Closing Price}}{\text{Earnings Per Share (EPS)}}$$

- **Return on Equity (ROE):** Implied (from the Income Statement and Balance Sheet).

$$ROE = \frac{\text{Net Income}}{\text{Total Shareholders' Equity}}$$

- **Debt-to-Equity (D/E):** Implied (from the Balance Sheet). These features are implied on a quarterly/annual basis.

$$D/E = \frac{\text{Total Liabilities}}{\text{Total Shareholders' Equity}}$$

B. Technical Time Series Features

· **The LSTM Model:** Features that assess **Timing and Momentum**. These are features that create derived indicators from **Relative Strength Index (RSI)**, **Moving Average Convergence Divergence (MACD)**, **Bollinger Bands (BB)** and **Exponential moving average (EMA)/Moving Averages (MA)**. Input window 60 **trading days** 840 numerical values for each prediction day.

4.3. Pre-processing and Matching

1. **Matching:** Financial values match until the next quarterly announcement since it's more fundamental data (less frequent than daily trading data).
2. **Normalization:** Min-Max-Scaler applies the min-max all features valued [0, 1]
3. **Y Variable Definition:** A signal is successful (Y=1) if a **+5% return** is achieved before a **-2% loss** (stop-loss) within 30 trading days after the buy signal is given.

4.4. Stacking Ensemble Architecture Evaluation

The predicted outputs of the two base learners - the LSTM network (relating to assessing **Timing**) and the XGBoost model (related to assessing **Value**) - will be the probability score and $P_{XGBoost}(\text{Value})$. Probability is defined as to what extent the binary target variable $Y=1$ will appear. These will be presented as meta-features (a higher-level input) for the logistic regression meta-learner. The main role of this Meta-learner is optimization: it will assess what the joint thresholds, $T_{\{\text{Value}\}}$ and $T_{\{\text{Timing}\}}$ should be, which ultimately on the test set will maximize the Sharpe Ratio, meaning the final trading rule will ensure that any buy signal should only be issued if risk-adjusted return is guaranteed. Thus, the final actionable buy signal can be made when the two base models agree on a predicted high-probability occurrence that satisfies the conditional logic of the following two conditions.

6. References

- Al-Nefaie, A. H., & Aldhyani, T. H. H. (2022). Predicting Close Price in Emerging Saudi Stock Exchange: Time Series Models. *Electronics*, 11(21), 3443. <https://doi.org/10.3390/electronics11213443>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Capital Market Authority. (2021). Market conduct regulations. Retrieved from <https://cma.org.sa/en/Pages/MarketConductRegulations.aspx>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Daori, H. K., Al-Shaibani, M. A., Alsaif, S., Alahmari, A., & Alshahrani, M. (2024). Saudi market stock price predictions using deep learning algorithms and weighting Twitter sentiment analysis. *University of Ha'il-Journal of Science (UOHJS)*, 5(1).
- Daradkeh, M. K. (2022). A hybrid data analytics framework with sentiment convergence and multi-feature fusion for stock trend prediction. *Electronics*, 11(2), 250. <https://doi.org/10.3390/electronics11020250>
- Graham, B. (2006). *The intelligent investor: The definitive book on value investing*. HarperBusiness.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Khurayshi, A., & Alharbi, A. (2025). Saudi Arabia stock prices prediction using machine learning. *Advances and Applications in Statistics*, 92(6), 813–833.
- Yahoo Finance. (n.d.). Saudi stock market data (Tadawul). Retrieved from <https://finance.yahoo.com/markets/>