

AI-Driven Stock Market Forecasting Using Financial Statement Analysis

By

Marwa Othman Esaa

Midocean University

Advisor

Dr. Ashraf Elsayed

Abstract

Stock prediction has attracted considerable attention among investors, with recent interest focused on the use of machine learning techniques to enhance predictive accuracy. Numerous studies have proven the effectiveness of machine learning in forecasting stock market movement, either using technical, fundamental, or sentiment analysis.

This work introduces a practical tool for stock market investors by employing five machine learning models: Random Forest (RF), Feedforward Neural Network (FNN), Gated Recurrent Unit (GRU), Graph Attention Network (GAT), and Financial Graph Attention Network (FinGAT). The models are trained and tested using financial statements and historical stock data for 253 Saudi companies listed on the Tadawul All Share Index (TASI). By effectively analyzing financial statements and historical stock data, the proposed models aim to increase investor profits, reduce financial risk, and highlight the effectiveness of employing machine learning techniques based on financial ratios for stock prediction.

Introduction

In selecting the fiscal year-end for analysis, this study aligns itself with the Saudi stock market's (TASI) fiscal cycle to ensure uniformity in evaluating mid- to long-term directions of stocks. The TASI was chosen because Saudi-listed companies tend to synchronize annual reporting periods, and it facilitates cross-firm comparison and increases the reliability of time-series analysis. Alternatively, different fiscal year-ends for different companies might cause reporting lapses that would hinder accurate measurement of performance.

This research employs three primary stock analysis techniques:

Technical Analysis: Examines historical stock prices and trading volumes to identify patterns and forecast direction.

Fundamental Analysis: Evaluates a company's underlying value based on financial statements, profitability, management efficiency, and products.

Sentiment Analysis: Studies investor sentiment by reading news media, social media, and other public materials. (it is **recommended** that sentiment data be integrated into the dataset in future extensions of this research.)

Though highly volatile and exposed to exogenous factors such as economic signals, geopolitical crises, and psychological sentiments, share prices do not necessarily reflect a company's underlying worth. Fundamental analysis sheds greater light on a company's financial health and longer-term destiny.

Therefore, it is best adapted for forecasting stock activity quarter-by-quarter and year-by-year on a quarter-by-year basis.

It is important to note that a fiscal year (FY) refers to a specific start and end time chosen by a company based on its business nature and revenue cycle. While many companies align their fiscal year with the calendar year, US stock market companies are allowed to freely choose their fiscal year start date, which does not necessarily begin on 1 January. But, in Saudi Arabia, the fiscal year is the calendar year, 1 January to 31 December. Financial reports follow standardized release dates, with Q1 quarterly reports released before 15 May, Q2 before 11 August, Q3 before 11 November, and annual and Q4 financial reports before 31 March of the following year.

This paper is organized as follows: it begins with a problem specification, followed by motivations and objectives. It then reviews the relevant literature, presents the proposed methodologies, and describes the dataset that will be used to train and test the model.

Problem Specification

Accurate forecasting of stock price movements is one of the most challenging problems in finance due to the inherent volatility, non-linearity, and complexity of financial markets. Traditional methods of forecasting through statistical models and human inspection consistently miss the vast universe of internal and external drivers affecting stock prices such as firm performance, sentiment, economic indicators, and global events.

Equity investors within the Saudi equity market (Tadawul) require enhanced decision-making aids, especially in the case of mid- and long-term investment plans. Short-term signals can be obtained using technical and sentimental analysis, whereas fundamental analysis relies on financial reports provides a higher level of awareness regarding the company's intrinsic value and long-run potential. However, conventional fundamental analysis is limited by human bias, interpretation variation, and the growing volume of financial data. This requires the need for intelligent, automated methods that can review historical financial data and identify meaningful patterns.

Therefore, the underlying problem this paper is addressing is:

How can stock price change machine learning models be used to generate reliable and scalable forecasting for predictions of accurate stock price changes based on Saudi financial statement data?

To solve this, the study recommends developing and experimenting with a few machine learning models, Random Forest, FNN, GRU, GAT, and FinGAT trained on financial statements of 253 listed companies on the Tadawul All Share

Index (TASI) to decide on the optimum model for stock trend forecasting and providing actionable insights to investors.

Motivations

Stock market prediction has long been a vital area of interest to investors, financial analysts, and researchers alike. With the rapid increase in volume and sophistication of financial data, traditional forecasting methods have failed to cope with the non-linear and dynamic nature of modern markets especially in developing markets like Saudi Arabia's Tadawul All Share Index (TASI). Machine learning (ML) has recently proved successful in addressing such challenges by offering tools that can analyze enormous data sets, identify hidden patterns, and offer improved predictions. However, most of the research work done has been on developed markets, with none of them employing advanced ML models on Saudi-listed companies based on fundamental financial data. More appealing to this research is also the consistency in fiscal reporting in the Saudi market, where a calendar fiscal year is adopted consistently by all listed companies. Such alignment enables more effective time-series analysis and model training, which is usually challenging across markets where the reporting period does not align. Additionally, investors now increasingly seek evidence-based solutions for enhanced decision-making and reduced risk. Based on financial ratios and machine learning algorithms, this study attempts to bridge the divide between traditional financial analysis and current AI-driven forecasting and present a handy tool for evidence-based more effective investment in the Saudi stock market.

Objectives

The primary objective of this study is to develop and evaluate machine learning models to forecast stock price movement in the Saudi stock exchange using financial statement information. The research aims specifically:

To apply and contrast different machine learning algorithms, the Random Forest (RF), Feedforward Neural Network (FNN), Gated Recurrent Unit (GRU), Graph Attention Network (GAT), and Financial Graph Attention Network (FinGAT), to predict stock price variations.

Leverage financial ratios and historical stock performance data of 253 Tadawul All Share Index (TASI) listed firms as the groundwork for fundamental analysis.

Demonstrate the potential of machine learning techniques to enhance mid- to long-term investment decision-making in terms of reducing prediction error and identifying high-performing stocks.

Literature Review

In recent years, numerous studies have explored the application of machine learning techniques in stock prediction. Gupta, H., & Jaiswal, A. [1] in 2024 introduced a study on stock forecasting using deep learning and statistical models. The survey motive is to check various deep learning and statistical model techniques for stock price forecasting that are Moving Averages, ARIMA which are statistical techniques and LSTM, RNN, CNN, and FULL CNN which are deep learning models. The result of all the models is different and the best result comes from the LSTM model and we have also seen how Full CNN outweighs the pre-process CNN. This model helps the investors to know when to sell and buy the stock prices. As is also known how the stock prices are affected by the climatic or geographical changes, company profit or loss, change in geographical seasons, loss in the region, inflation, competition in the market etc. Additionally, Tsai, P.-F., Gao, C.-H., & Yuan, S.-M. [2] in 2023 delved into deep learning models for this purpose, they introduced a paper about, Stock selection using machine learning based on financial ratios, found that successful utilization of RF, FNN, GRU, and FinGAT models holds significant managerial implications for enhancing portfolio management strategies. On the other hand, Kunathananon, T. [3] in 2022 introduced *Estimating stock price based on information from financial statement using machine learning approaches*. This study introduces a tool for stock market investors and institutions constructed from Long Short-Term Memory (LSTM) for predicting stock prices. By effectively analyzing financial statements and stock market data, the results indicate that LSTM can effectively capture complex relationships in the data and predict stock prices. This research highlights the potential of LSTM as a valuable and innovative tool for investors and institutions in the stock market. In a historical context, the works of Heo, J., & Yang, J. Y. [5] in 2016 this study about stock price prediction based on financial statements using SVM. Han, S., & Chen, R.-C. [6] in 2007. The study used SVM with financial statement analysis for prediction of stocks.

Methodology

This study will use a deep learning approach to stock price movement prediction in the Saudi stock market (Tadawul All Share Index – TASI) using analysis of financial statements. The approach is intended to extract important financial features and apply time-series modeling techniques to predict future stock price movements.

The proposed Deep Learning models: [2]

Random Forest (RF)

In the realm of stock selection, Random Forest employs a strategic approach. It leverages a random subset of training data and input characteristics to craft each tree within the set. This approach serves the dual purpose of mitigating overfitting concerns and bolstering the model's capacity to generalize effectively when confronted with unseen data. By incorporating the practice of bootstrap aggregation (bagging), the algorithm further augments tree diversity and curtail variance.

Feedforward Neural Network (FNN)

FNN plays an invaluable role in stock selection by adeptly modeling and scrutinizing non-linear associations between inputs and outputs. Leveraging the backpropagation algorithm, the FNN can be trained to enhance its performance and elevate the precision of stock predictions and investment decisions.

Gated Recurrent Unit (GRU)

GRU, with its adept handling of sequential data, proves highly effective in analyzing historical stock data and uncovering influential patterns and trends that can impact future stock performance. As a result, GRU assumes a pivotal role in stock prediction models, contributing to the precision of stock selection for investment purposes.

Graph Attention Network (GAT)

GAT is a specific subtype of graph neural network (GNN) renowned for its proficiency in dissecting the intricate relationships prevalent in stock markets. The financial domain has witnessed a growing number of investigations delving into the utilization of graph neural networks (GNNs) due to their adeptness at comprehending the intricate linkages between stocks. This research has neatly outlined the prevailing structural patterns of GNNs as applied in finance.

Financial Graph Attention Network (FinGAT)

FinGAT stands as an advanced neural network meticulously designed for the precise prediction of stock returns. It effectively harnesses the power of graph attention mechanisms to grasp the intricate relationships between diverse stocks. The system comprises three pivotal components: the Gated Recurrent Unit (GRU) for sequential learning, the intra-sector relationship within the Graph

Attention Network (GAT), and the inter-sector relationship of GAT, the synergy of GAT and GRU is the defining attribute of FinGAT, enabling it efficiently to comprehend and leverage the relationships both within and among sectors of stocks.

Data Description

Historical financial statement data and market prices were obtained for listed companies on the Tadawul Stock Exchange, the dataset includes data from 253 publicly traded companies across various sectors, including banking, energy, industrials, healthcare, and real estate.

Source: Tadawul and company reports, Yahoo Finance, SAMA and GASTAT

Financial Ratios of 18 Ratios as Attributes:

There are 18 financial ratios as input for models, which are the same as those used in research [2]. The 18 financial ratios are listed, and the calculation formula is given in the following table.

Type	Ratio	Calculation
Liquidity Ratios	Current	$\frac{\text{Current assets}}{\text{Current liabilities}}$
Leverage Ratios	Debt to Equality	$\frac{\text{Long Term Debt}}{\text{Shareholders' Equality}}$
	Debt to Capital	$\frac{\text{Long Term Debt}}{\text{Shareholders' Equality} + \text{Long Term Debt}}$
Asset Efficiency Ratios	Asset Turnover	$\frac{\text{Sales}}{\text{Total Assets}}$
	Inventory Turnover	$\frac{\text{Sales}}{\text{Inventories}}$
	Receivable Turnover	$\frac{\text{Net Credit Sales}}{\text{Receivables}}$
	Days Sales in Receivable	$\frac{\text{Receivables}}{\text{Net Credit Sales}} \times 365$
Market Value Ratios	Book Value per Share	$\frac{\text{Shareholders' Equality}}{\text{Shares Outstanding}}$
Probability Ratios	Gross Margin	$\frac{\text{Sales} - \text{Cost of Goods Sold}}{\text{Investment Base}}$
	Operating Margin	$\frac{\text{Operating Income Before Interest and Taxes}}{\text{Revenue}}$
	Pre-tax profit margin	$\frac{\text{Sales} - \text{Cost of Goods Sold}}{\text{Investment Base}}$
	Net Profit Margin	$\frac{\text{Net Operating Income}}{\text{Revenue}}$
	Return on Equality	$\frac{\text{Net Operating Income}}{\text{Shareholder's Equality}}$
	Return on Tangible Equality	$\frac{\text{Net Operating Income}}{\text{Tangible Shareholder's Equality}}$
	Return on assets	$\frac{\text{Net Operating Income}}{\text{Total Assets}}$
	Return on investment	$\frac{\text{Investment Gain}}{\text{Investment Base}}$
	Operating Cash Flow Per Share	$\frac{\text{Operating Cash Flow}}{\text{Shares Outstanding}}$
	Free Cash Flow per Share	$\frac{\text{Operating Cash Flow} - \text{Capital Expenditures}}{\text{Shares Outstanding}}$

There are five main types of financial ratios:

Liquidity ratios: Liquidity ratios measure a company's ability to pay off its short-term debts.

Leverage ratios: Leverage ratios measure the amount of debt a company has relative to its assets or equity. These ratios are often used by investors and creditors to assess the riskiness of a company's operations and its ability to meet long-term debt obligations.

Asset efficiency ratios: Asset efficiency ratios measure how effectively a company uses and manages its assets to generate revenue.

Market value ratios: Market value ratios are used to evaluate a company's stock price in relation to its earnings, sales, and book value.

Time Period

Data covers the period from 2015 to 2023, using quarterly financial reports.

Evaluation Metrics

To assess the performance of the stock market prediction models developed in this study, a set of **standard evaluation metrics** for regression and classification tasks are used. These metrics help determine the accuracy, reliability, and robustness of each model when forecasting stock price movements based on financial statement data.

Mean Absolute Error (MAE)

MAE measures the average magnitude of the errors between predicted and actual stock prices.

Root Mean Squared Error (RMSE)

RMSE is the square root of the average of squared differences between predicted and actual values.

Timeline

Week	Planned Activity
Week 1-3	Data Preprocessing (financial statement data, macroeconomic indicators, Clean and structure data, align financial and price timelines)
Week 4-7	Model Development
Week 8 - 9	Evaluation
Week10	Final Report & Submission

References

1. Gupta, H., & Jaiswal, A. (2024). A study on stock forecasting using deep learning and statistical models. arXiv. <https://arxiv.org/pdf/2402.06689v1>
2. López Gil, G., Duhamel-Seblin, P., & McCarren, A. (2024). *An evaluation of deep learning models for stock market trend prediction*. Dublin City University, School of Computing. <https://arxiv.org/abs/2408.12408>
3. Tsai, P.-F., Gao, C.-H., & Yuan, S.-M. (2023). Stock selection using machine learning based on financial ratios. *Mathematics*, MDPI, 4758. <https://doi.org/10.3390/math11234758>
4. Kunathananon, T. (2022). *Estimating stock price based on information from financial statement using machine learning approach* (Master's thesis, Chulalongkorn University). Chula Digital Collections. Retrieved from <https://digital.car.chula.ac.th/chulaetd/5818>
5. Heo, J., & Yang, J. Y. (2016). Stock price prediction based on financial statements using SVM. *International Journal of Hybrid Information Technology*, 9(2), 57-66. <https://doi.org/10.14257/ijhit.2016.9.2.05>
6. Han, S., & Chen, R.-C. (2007). Using SVM with financial statement analysis for prediction of stocks. *Communications of the IIMA*, 7(4), Article 8. <https://doi.org/10.58729/1941-6687.1059>