Capstone Project Literature Review

Ri-on Kim, Manuelito Bagasina, Youngjun Jhee

January 2025

1 Introduction

1.1 Background

Stock price forecasting has been a challenging job because of the intricate interaction of economic, financial, and psychological factors. The conventional financial theories, namely the **Efficient Market Hypothesis** (EMH) and **Random Walk Theory**, opine that stock prices undergo a stochastic process and are almost impossible to forecast[17]. However, with advances in **Machine Learning** (ML) and access to large datasets, researchers have come up with intricate predictive models that utilize or combine several sources of information, such as historical stock prices, macroeconomic variables, firm-specific financial variables, and market sentiment.

1.2 Purpose of the Review

The purpose of this literature review is to summarize current stock market price prediction research based on machine learning and data science in order to establish a solid basis for the DAT 490 capstone project. The review specifically seeks to:

- **Understand Existing Approaches:**

  - Study conventional stock forecasting approaches, including technical and fundamental analysis, in order to determine their limitations.

  - Describe contemporary machine learning methodologies, such as supervised learning (neural networks, regression) and unsupervised learning (clustering) and ensemble methods.

- **Identify Key Factors Affecting Stock Prices:**

  - Assess the influence of macro drivers (GDP growth, interest rates, CPI), micro drivers (P/E ratio, earnings per share) and historical stock prices on stock price movement.

  - Explain how sentiment analysis on news feeds and social media aids in enhancing predictive models.

- **Evaluate the Effectiveness of Different Models:**

  - Compare a range of different machine learning algorithms for stock price prediction.

  - Identify which features are most important for making accurate predictions.

- **Highlight Research Gaps and Challenges:**

  - Discuss challenges like data quality, market efficiency, feature selection, and ethical issues in AI-based trading.

  - Suggest areas of future research, e.g., reinforcement learning, alternative data sources, and explainable AI.

- **Inform the Project's Methodology:**

  - Direct the choice of data sources and preprocessing methods.

  - Clarify the research question by knowing the strengths and weaknesses of current work.

  - Help choose the most appropriate machine learning methods for stock market prediction.

2 Existing Approach

2.1 Data Collection and Processing

2.1.1 Microeconomic Indicators

Microeconomic indicators mirror the financial well-being and efficiency of individual firms. These indicators are essential for investors who need to determine stock valuation and make relevant investment choices.

Key Microeconomic Indicators:

- **Earnings Per Share (EPS)** - EPS gauges the ratio of a firm's profit distributed to each outstanding share of common stock. Increased EPS typically indicates increased profitability and positively influences stock prices [1].
- **Price-to-Earnings (P/E) Ratio** - This gauges a company's current stock price in relation to earnings per share. It assists investors in deciding if a stock is overvalued or undervalued [20].
- **Debt-to-Equity Ratio** - This approximates a company's financial leverage by utilizing total liabilities to shareholders' equity. A higher ratio signifies greater financial risk, which affects the volatility of stock [14].
- **Return on Assets (ROA)** - ROA measures how effectively a company puts assets to use in order to generate profit. Firms with greater ROA have greater investor confidence [40].
- **Return on Equity (ROE)** - ROE measures the ability of a firm to generate profit from the shareholders' investment. Higher ROE tends to attract investors who seek efficient use of capital [10].

2.1.2 Macroeconomic Indicators

Macroeconomic indicators are more representative of overall economic trends and have a significant influence on the stock market's performance.

Key Macroeconomic Indicators:

- **Gross Domestic Product (GDP)** - GDP growth would typically align with rising stock prices as it is indicative of economic growth and higher corporate earnings [40].
- **Inflation Rate (CPI)** - Moderate inflation is favorable for economic growth, but runaway inflation erodes purchasing power and affects corporate profitability and hence stock prices [24].
- **Interest Rates** - Interest rates are managed by the central banks to manage inflation and economic growth. An increase in interest rates will typically result in a decline in corporate investment and share prices [30].
- **Unemployment Rate** - Increasing unemployment is typically a sign of economic turmoil, decreasing consumer spending and corporate revenues, resulting in falling share prices [29].
- **Exchange Rates** - Volatility in exchange rates impacts global companies' revenues. A depreciating home currency benefits exporters but makes imports expensive, influencing overall market performance [21].

2.1.3 Stock price

The previous two sections look at the intrinsic value of a stock based on things like the economy, performance of the company, or perhaps political climate to predict the value of a stock. This section focuses on looking at a stock's historical data to make predictions. This section will focus on the statistics generated by past prices and volumes of stocks, akin to the 2015 Patel and Shah paper [39]. The view from this section is that raw numbers are more logical than other factors, and that if data is preprocessed properly and an appropriate algorithm is used, then stock price prediction accuracy increases.

A popular variable/technical factor used in this field is the **open-high-low-close** chart (OHLC). This is a chart that shows the opening price, the highest and lowest prices reached, and the closing price of a stock [8]. The 2018 Jagwani paper used stock prices from January 2000 to January 2018 that they gathered from Yahoo Finance. They then used time series analysis, analysis that uses data collected over a series of time, on the data to make predictions. [26].

The data collected is certainly valuable, but the key difference between different studies comes down to statistics/machine learning methods and how data should be preprocessed. The early days of this field focused on publicly available information regarding a stock and its stock returns. This simple and linear method of thinking is argued to be inaccurate as stock price and the available information aren't in a linear relationship. That is why a variety of different machine learning methods must be used. In the 2015 Patel paper, different Machine Learning methods were used and gave out these results: "The accuracy of 86.69%, 89.33%, 89.98% and 90.19% is achieved by ANN, SVM, random forest and naive-Bayes (Multivariate Bernoulli Process) respectively." [39] . Since this study will focus on multiple factors, it is crucial that the right method be used to get the most accurate model.

The key problem then would be deciding which model to use and which data preprocessing method should be used. Another key consideration would be how fine-tuning each model would be resources-straining as well. In the Enke paper, it was mentioned how training the models was computationally expensive [16]. Common machine learning methods used were artificial neural network, support vector machine and fuzzy logic. There is also a growing interest in using **Deep Learning** as more studies are conducted on its effects, so perhaps this paper can also research on this method to expand the field. [39].

2.1.4 Sentiment Analysis

Twitter is a social media platform that allows users to post their own thoughts or reply to others. 140 million tweets are posted everyday, making the platform a goldmine for data [27]. This data can be utilized to make predictions on a stock based on people's thoughts on that particular company, and the method to do so is called sentiment analysis. Sentiment analysis can be described as describing a text's opinion as positive, neutral, or negative [37]. The field has evolved over time, with new innovations like: more efficient data collection, better cleaning/filtering, detailed classification of emotions, and better machine learning methods.

Key Studies done on Sentiment Analysis

- **Bollen 2011 -** This study used 9.8 million tweets from 2.7 million users (Feb–Dec 2008) for its dataset. It filtered out spam and URL's from its tweets and used mood-related keywords like "I feel" or "makes me". It analyzed the **Dow Jones Industrial Average (DJIA)** as its basis to see if Twitter mood helped analyze the stock market's performance.

  It used two sentiment analysis tools, **Opinion Finder** (OF) and **General Profile of Mood States** (GPOMS). OF creates a dictionary of positive and negative terms and creates a binary classification of tweets based on these terms' presence. GPOMS expanded the moods into more types, such as: Calm, Alert, Sure, Vital, Kind, and Happy; and computed for each of these segments and then accumulated over all segments.

  To validate the data, a time-series was pitted against USA events, like Thanksgiving or Election day. They also used **Granger Causality**, which tests predictive relationships between mood dimensions and DJIA. They also used **Self-Organizing Fuzzy Neural Network** (SOFNN), which predics DJIA using historical prices and mood data.

  The key results were that there were more nuanced moods when it came to the same events, like how pre-election has moods of anxiety and excitement while post-election can have moods of joy and depression. There was a 87.6% accuracy when it came to the DJIA direction. This study eventually proves that real-time social media sentiment can help predict stock market movement.

There is still room for improvement. This study only focuses on the USA and English speakers. Events like the 2008 bank bailout caused DJIA spikes unreflected in mood data. The binary classification caused by OF was cited to be less accurate compared to the more nuanced GPOMS. [3].

- **Pagolu 2016** - This study used 250,000 tweets (2015–2016) that mention "Microsoft", "$MSFT", or "Microsoft". They used the closing price from Yahoo Finance as the variable to observe. The main contribution of this study is analyzing the performance of two textual representations, **N-gram** and **Word2vec**, to analyze the sentiment from tweets. N-grams break text into word chunks (e.g., three consecutive words) and use binary vectors to indicate their presence (either a 1 or 0), offering simplicity but limited context understanding. Word2vec uses 300-dimensional vectors to capture semantic relationships (words like "happy" or "joyful get grouped near each other), which allows a more nuanced sentiment to be made.

  The study also had a similar success where they proved that a positive sentiment on a company tends to correlate with an increase in stock price. It also demonstrated that methods that capture more nuance from a tweet do better than simpler methods, in the case of Word2vec being better than N-grams. The study also shows the method of sentiment analysis, from getting the data, preprocessing the data, giving data to the sentiment analyzer, then finishing with a correlation analyzer. Unfortunately, this study does cite some shortcomings. Firstly, sarcasm would make a supposedly negative tweet into something positive. Secondly, not every investor voices their thoughts on Twitter. Lastly, the study suggested that more data/tweets be used to gain a more accurate prediction [37].

- **Khedr 2017** - This paper is quite similar to the Pagolu paper, but it uses News Headline articles along with historical data (OHLC) of a stock. It focused on Yahoo, Microsoft, and Facebook for its paper The paper also gathered 3 types of news for its data: news relevant to the market, company news, and financial reports that were published by financial experts about stocks. It classified the news data as either positive or negative.

  It claims to have a higher accuracy in its models through its use of **Naïve Bayes** algorithm along with **NLP** techniques and T**f-idf**. The paper also studied Microsoft, along with other companies, and discovered that its Microsoft prediction was less accurate due to Microsoft not having as many news articles written about it, leading to less data. The key point in this study is this: the models had a higher accuracy if they combined both historical OHLC data along with news sentiment analysis [31].

- **Lee 2023** - This paper came attached with its own dataset of financial sentiment/emotion classification and stock market time series prediction, which can be found here:
  https://github.com/adlnlp/StockEmotions

  The paper has many new innovations from previous papers. It has a lexicon for emojis, uses **Natural Language Processing** (NLP), and uses 9 machine learning methods. It collected tweets from Jan 2020 - Dec 2020, uses two sentiments bullish (believes the stock will go up) and bearish (believes the stock will go down), and has 12 emotions (ambiguous, amusement, anger, anxiety, belief, confusion, depression, disgust, excitement, optimism, panic, surprise). It uses StockEmotions from StockTwits, which is a database of 10,000 English comments. Its uniqueness stems from investor sentiment classes, fine-grained emotions, emojis, and time series data. All these new sources of data are given to 9 Machine Learning models, and they all vary in terms of accuracy. The models used were:

  - Logistic Regression,
  - Naive Bayes SVM

- standard neural network
- GRU
- Bi-GRU
- pre-trained language models
- DistilBERT
- BERTbase
- RoBERTabase

The study has its own limits as well. It cites potential bias in including offensive language, user base biases, and annotator bias. StockTwit is also a platform known for speculating on "meme' stocks, so tweets collected might contain toxic content, offensive language, and sarcasm. [33].

2.2 Machine Learning and Modeling Method

2.2.1 Traditional Statistical Models

Prior to ML, stock market forecasting depended on time-series statistical models. They are still relevant in financial analysis.

- **Autoregressive Integrated Moving Average (ARIMA)**: ARIMA models learn historical price patterns by capturing historical values [6]. Although adequate for short-term predictions, they fail to handle the non-linearity of the financial markets. Extensions such as SARIMA (Seasonal ARIMA) enhance performance for seasonal trends in stocks [25].
- **Vector Autoregression (VAR)**: VAR models examine interdependencies among several time-series variables, for instance, stock prices, interest rates, and GDP [44]. They can be utilized to represent macroeconomic effects on stock prices.
- **Generalized Autoregressive Conditional Heteroskedasticity (GARCH)**: GARCH models are developed for volatility modeling and find extensive applications in risk measurement and portfolio management [15]. GARCH models are efficient in market volatility forecasting and find particular applications in derivative pricing [5].

2.2.2 Machine Learning Approaches

ML algorithms boost forecasting in the financial market via feature selection automation, complex pattern extraction, and processing of high-dimensional datasets.

- **Supervised Learning Methods**
  - **Regression Models**:
    - Linear Regression: Simple and interpretative model but takes into consideration only linear relationships between factors of stock price [22].
    - Support Vector Regression (SVR): Can capture non-linear relationships better than linear regression [9].
  - **Decision Trees and Ensemble Methods**:
    - Random Forests: Uses multiple decision trees to improve prediction accuracy and mitigate overfitting [7].
    - Gradient Boosting (GBM, XGBoost, LightGBM): Most common in financial forecasting for feature importance and high accuracy in prediction [11].
  - **Neural Networks and Deep Learning**:
    - Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM): Effective for time-series forecasting [23].

        * Transformer Models: Outperform LSTMs in handling long-term dependencies in financial data [45].

- **Unsupervised Learning Methods**
    - K-Means Clustering: Classified stocks with regards to price movement similarity [19].
    - Principal Component Analysis (PCA): Dimensionality reduction in stocks, improving computational efficiency [28].
    - Autoencoders: Used for anomalous price movement behavior in stocks [36].

2.2.3 Hybrid and Multi-Modal Approaches

Many researchers combine multiple methods to improve prediction accuracy.

- **Hybrid Models**
    - LSTMs for prediction in a times-series, with feature consideration added for refinement through Random Forest or XGBoost [13].
    - CNN + LSTM: Convolutional Neural Networks (CNNs) extract graphical structures in plots of stocks, with LSTMs processing sequential trends in price for efficient forecasting in stocks [43].
- **Sentiment Analysis and Stock Price Prediction**
    - Twitter-based Sentiment Analysis: Analysis of social media for tweets' sentiment has been successful in financial prediction [4].
    - Financial News Analysis: NLP algorithms scan through headlines, articles, and expert reports for an impact analysis in stocks [41].
    - Transformer-based NLP Models: FinBERT, a financial sentiment classification model, a fine-tuned model for improvement in traditional text analysis [2].

Machine learning and deep learning have become ever more capable at predicting prices in financial markets through blending together many sources of information. Hybrid techniques and deep architectures such as LSTMs and Transformers yield state-of-the-art performance, but overfitting, interpretability, and data quality will become important areas for future work.

3 Research Gaps

3.1 Limitations of Existing Studies

3.1.1 Data Collection

- Microeconomic Indicators: Gathering microeconomic data is also plagued with several issues in relation to predicting the stock market. Although financial statements and balance sheets of listed companies are easily accessible, the reliability of financial ratios like Earnings Per Share (EPS), Price-to-Earnings (P/E) Ratio, and Return on Assets (ROA) is reliant on efficient accounting procedures [1]. Furthermore, quarterly financial reports are generally published, and real-time prediction is not possible. Furthermore, variation in financial reporting standards between markets creates inconsistency in data quality and cross-market analysis becomes cumbersome [20]. There is some evidence to indicate that real-time transactional data, for example, insider trading activity or investor sentiment at the firm level, would be more predictive, but data of this nature are frequently not available due to restrictions by regulatory authorities [14].

  A second issue is that company-reported data is susceptible to manipulation, which can represent skewed actual firm performance [40]. In addition, economic shocks, i.e., financial crises or policy changes, impact firm-level financial health in ways that historical microeconomic indicators cannot fully capture [10]. Additional research is needed to integrate other firm-level data sources, i.e., supply chain disruptions and real-time transaction-level metrics, to enhance the predictive power of microeconomic indicators.

- Macroeconomic Indicators: Macroeconomic data gathering has its own constraints that affect stock market prediction. In contrast to stock price information, which can be retrieved at high frequency (e.g., daily, minute-by-minute), macroeconomic variables like GDP growth, interest rates, and unemployment rates are typically accessible on a quarterly or monthly basis, which limits their usefulness in high-frequency trading models [24]. Secondly, economic news is revised periodically, so macroeconomic past data can be untrustworthy for long-term forecasting [30].

  A second challenge is that macroeconomic variables impact one another in complex, interrelated ways that are difficult to untangle. For instance, a rise in interest rates might be a signal of economic strength (combating inflation) or economic weakening (constricting credit availability), with opposing implications for share prices [29]. Geopolitical risks, trade policy, and currency markets also introduce noise into macroeconomic models, making any predictive relationships difficult to nail down [21].

  Whereas most studies use conventional economic variables, recent research indicates that real-time macroeconomic indicators, including satellite metrics of economic activity and non-conventional financial transaction data, can provide more timely leads in predicting stock prices [12]. Future research should try to include such unconventional macroeconomic data sources for enhancing forecasting accuracy.

- Stock Price: The data collection of the stock price is fairly simple, one only has to weekly/monthly/quarterly data on open, high, low, close and adjusted close prices of stocks from the Yahoo Finance API. Historical data is very simple to train a machine learning model, but its main shortcoming is that all of the data is based purely on past data. It doesn't take into account financial news that happens on a daily basis, which can significantly change a stock's trend in value. [26]

- Sentiment Analysis:

  The problems cited by many of the studies are filtering and cleaning the data. For news articles, some studies didn't take into account sarcasm and other relevant contexts that can alter the sentiment of the headline, which might use "positive" words [31]. In every study that web-scraped tweets, several filters had to be used to sift out things like spam, promotional materials, URLs [35]. The Pagolu paper mentioned how not every investor voices their opinions on Twitter, so there may be issues wherein something financially terrible happened to a company but nobody on Twitter talked about it [37]. It must be mentioned that many studies focus on English, leaving out data for other languages. It must also be mentioned that some studies suggest future work to gather an even larger dataset than their study performed, saying that it can lead to a more accurate prediction model.

3.1.2 Machine Learning and Modeling Method

Despite the improvement in machine learning (ML) algorithms for predicting stock price, a variety of problems still prevail.

- Feature Selection and Model Transparency: One of the biggest vulnerabilities of ML financial forecasting models is feature selection. Most studies utilize technical indicators, price history, and macroeconomics, but nobody can say with confidence what feature combination will perform best. Feature engineering is heuristic, and feature selection algorithms such as recursive feature elimination (RFE) and SHAP values have not been researched enough in forecasting stock markets[32]. To add, model interpretability is not high in deep learning algorithms such as LSTMs and transformers, and model selection cannot be understood, a significant drawback in a financial environment in which transparency is key[23].

- Overfitting and Over-Generalization: Stock price information is noised and non-stationary, and deep learning algorithms have a high vulnerability to over-fitting. Most studies have high training accuracy but lack generalizability in new, unseen samples through information spillover or over-dependency in preceding trends. Methods such as dropout, batch normalization, and Bayesian neural networks have been mooted, but performance varies with datasets[13].

- Hybrid Model Performance: Some studies have mooted combining two or three ML approaches, such as LSTMs and XGBoost, and CNNs with an attention mechanism, in an attempt to maximize prediction accuracy. Hybrid approaches, however, have high hyperparameter tuning requirements and high computational requirements. To add, performance is sensitive to information added in, and added complexity will not necessarily yield a proportionate improvement in prediction accuracy in most cases[11].

- Alternative Modeling Techniques: While deep algorithms have dominated in modern studies, algorithms including reinforcement learning (RL) and graph neural networks (GNNs) have been effective in financial workloads too. RL algorithms can adapt trading methods dynamically, and GNNs can utilize relational information including companies' relations and intermarket relations. Nonetheless, such algorithms have yet to be implemented, and any application in forecasting stocks in real life will have to undergo additional testing and analysis[33].

While machine learning advanced stock market prediction accuracy, there are still many limitations in this field. First, the merging and synchronization of data can be seen with macroeconomic indicators (monthly/quarterly) struggling to synchronize with daily stock prices and real-time sentiment data. Second, there is also the struggle with sentiment analysis not reflecting the mood of the market accurately, as nuance can be lost due to the complexity of language. Thirdly, feature selection will be the most difficult in this particular project. This project is going to combine so many features that finding the right combination itself would take time.

3.2 Future Directions

- Enhanced Feature Selection and Explainability: Feature selection is a key yet under-researched area of stock price prediction. Traditional approaches such as moving averages, volumes, and macroeconomics have been utilized in studies, but newer techniques such as SHAP values, deep learning's attention mechanism, and feature selection through reinforcement learning could make prediction models more interpretative. Future studies can work towards creating techniques for dynamically selecting key features according to changing market scenarios, minimizing noise, and enhancing prediction accuracy.

- Hybrid and Multi-Modal Approaches: Hybrid approaches such as combining LSTM and CNN have proven effective, but in practice, most such approaches have not gained widespread use. Integration of disparate modalities such as text (news sentiment analysis), images (patterns in a stock chart), and numerical values (tech analysis factors) in transformer architectures such as FinBERT and GPT-based financial frameworks could make prediction even more accurate. Research can work towards optimizing multi-modal architectures with efficient computations.

- Reinforcement Learning for Adaptive Trading Strategies Stock price prediction models predominantly rely on static approaches and lack adaptability according to changing scenarios in the marketplace. Reinforcement learning (RL) techniques, with continuous feedback with the marketplace for developing ideal trading strategies, can make them adaptable and effective in real-life scenarios. Research studies in the future can work towards utilizing RL-based agents for balancing between exploiting successful strategies and searching for new ones in turbulent markets.

- Overcoming Data Quality and Bias in Sentiment Analysis: Sentiment analysis models suffer from language, user, and platform biases in most studies. Most studies use tweets in the

English language and financial news and exclude tweets in any language and any cultural environment. Sarcasm and irony cause significant difficulty in sentiment classification, and future studies can attempt to use techniques such as cross-lingual sentiment models and emotion-aware NLP algorithms in an attempt to mitigate such biases.

- Model generalizability and robustness: Many deep architectures generalize well in training but suffer when generalizing in real financial markets. Overfitting is a significant challenge, especially with architectures such as LSTMs and Transformers, whose complex architectures make them susceptible to overfitting. Techniques such as meta-learning, domain adaptation, and Bayesian deep learning can strive towards developing robust models with high performance in any form of market environment and any economy cycle.

4 Conclusion

Stock price prediction in the stock market is an important and challenging field of financial study. This literature review discussed how machine learning models tap into past stock prices, microeconomic predictors, macroeconomic variables, and sentiment analysis in their endeavor to enhance forecasting capability. Classical financial theories propose that stock prices follow a stochastic process, and it is hard to forecast them precisely [18]. Nevertheless, advanced data science and machine learning tools have introduced new models that find hidden patterns in financial data [38].

Our review brings out the importance of taking microeconomic metrics like Earnings Per Share (EPS), Price-to-Earnings (P/E) Ratio, and Return on Assets (ROA) into account while deciding on firm valuation [1]. On the other hand, macroeconomic metrics like GDP growth, inflation, and interest rates tell us about general economic trends that influence stock prices [24]. In spite of these advances, there remain limitations. The majority of the research does not include real-time economic and financial information [21], sentiment analysis can be biased and open to misinterpretation [3], and choosing the best set of predictive features remains challenging [30].

Future directions include the application of different data points like satellite-based economic indicators, high-frequency trading data, and more advanced forms of sentiment analysis [42]. Deep learning models coupled with classical statistical methods can also take stock price prediction further [12]. Additionally, the resolution of data quality challenges, explainability of AI-based models, and market volatility will be important in propelling this field forward [29].

Finally, as artificial intelligence and machine learning continue to develop, the potential to forecast stock price movement more accurately will be enhanced, providing investors, policymakers, and financial analysts with useful information. Nonetheless, researchers need to continue being mindful of the limitations and ethical considerations of AI-driven financial forecasting in order to make informed and responsible decisions [34].

References

[1] Emre Alper and Hasan Kara. The effects of microeconomic factors on the stock market: A panel for istanbul stock exchange. *Theoretical and Applied Economics*, 24(1):103–114, 2017.

[2] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*, 2019.

[3] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[4] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[5] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[6] George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control.* Holden-Day, 1976.

[7] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[8] Thomas N Bulkowski. *Getting Started in Chart Patterns.* John Wiley & Sons, 2014.

[9] Longbing Cao and Francis Tay. Support vector machine with adaptive parameters in financial time-series forecasting. *IEEE Transactions on Neural Networks*, 14(6):1506–1518, 2003.

[10] Nai-Fu Chen, Richard Roll, and Stephen A. Ross. Economic forces and the stock market. *Journal of Business*, 59(3):383–403, 1986.

[11] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[12] Yi Chen and Ming Li. Economic forces in stock returns. *arXiv preprint arXiv:2401.04132*, 2024.

[13] Eunsuk Chong, Choong-Yong Han, and Frank C. Park. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83:187–205, 2017.

[14] Kaushik Dalal. The relationship between the stock market and economic indicators: Evidence from singapore and hong kong. *Bryant Economic Research Paper*, 1:1–22, 2018.

[15] Robert F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[16] David Enke and Suraphan Thawornwong. The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with applications*, 29(4):927–940, 2005.

[17] Eugene F Fama. Efficient capital markets. *Journal of finance*, 25(2):383–417, 1970.

[18] Eugene F. Fama. Stock returns, real activity, inflation, and money. *American Economic Review*, 71(4):545–565, 1981.

[19] Tak-Chung Fu. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164–181, 2011.

[20] Bülent Güngör and Enes Kaygın. Factors affecting stock price: Evidence from commercial banks in the developing market. *Journal of Governance and Regulation*, 4(4):372–377, 2015.

[21] Md. Shoriful Haque, Md. Sahin Amin, Jalal Miah, Dai Min Cao, and Abid Hossain Ahmed. Boosting stock price prediction with anticipated macro policy changes. *arXiv preprint arXiv:2311.06278*, 2023.

[22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning.* Springer, 2009.

[23] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[24] Andreas Humpe and Peter Macmillan. Can macroeconomic variables explain long-term stock market movements? a comparison of the us and japan. *Applied Financial Economics*, 19(2):111–119, 2009.

[25] Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice.* OTexts, 2018.

[26] Jai Jagwani, Manav Gupta, Hardik Sachdeva, and Alka Singhal. Stock price forecasting using data from yahoo finance and analysing seasonal and nonseasonal trend. In *2018 Second international conference on intelligent computing and control systems (ICICCS)*, pages 462–467. IEEE, 2018.

[27] Bernard J Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11):2169–2188, 2009.

[28] Ian T. Jolliffe. *Principal Component Analysis*. Springer, 2002.

[29] Takashi Kaizoji and Masatoshi Miyano. Stock market crash of 2008: An empirical study of the deviation of share prices from company fundamentals. *arXiv preprint arXiv:1607.03205*, 2016.

[30] Md. Faridul Hasan Khan and Md. Mamun Billah. Macroeconomic factors and stock exchange return: A statistical analysis. *arXiv preprint arXiv:2305.02229*, 2023.

[31] Ayman E Khedr, Nagwa Yaseen, et al. Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, 9(7):22, 2017.

[32] Mahinda Mailagaha Kumbure, Christoph Lohrmann, Pasi Luukka, and Jari Porras. Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, 197:116659, 2022.

[33] Jean Lee, Hoyoul Luis Youn, Josiah Poon, and Soyeon Caren Han. Stockemotions: Discover investor emotions for financial sentiment analysis and multivariate time series. *arXiv preprint arXiv:2301.09279*, 2023.

[34] Frederic S. Mishkin. Interest rates and stock market volatility. *Journal of Economic Perspectives*, 4(2):49–64, 1990.

[35] Anshul Mittal and Arpit Goel. Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*, 15:2352, 2012.

[36] Andrew Ng. Sparse autoencoder. *CS294A Lecture Notes, Stanford University*, 2011.

[37] Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. Sentiment analysis of twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, pages 1345–1350. IEEE, 2016.

[38] Janak Patel, Sahil Shah, Parth Thakkar, and Krishna Kotecha. Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42(4):2162–2172, 2015.

[39] Jigar Patel, Sahil Shah, Priyank Thakkar, and Ketan Kotecha. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications*, 42(1):259–268, 2015.

[40] Alfredo Peiró. Us stock market and macroeconomic factors. *Journal of Applied Business Research*, 32(1):325–340, 2016.

[41] Robert P. Schumaker and Hsinchun Chen. Textual analysis of stock market prediction using breaking financial news: The azfintext system. *ACM Transactions on Information Systems (TOIS)*, 27(2):1–19, 2009.

[42] Robert P. Schumaker and Hsinchun Chen. Textual analysis of stock market prediction using breaking financial news: The azfintext system. *ACM Transactions on Information Systems (TOIS)*, 27(2):1–19, 2009.

[43] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman. Stock price prediction using lstm, rnn and cnn-slstm models. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1643–1647, 2017.

[44] Christopher A. Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980.

[45] Ashish et al. Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.