iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: I Month of publication: January 2024

DOI: https://doi.org/10.22214/ijraset.2024.58167

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

Stock Price Prediction Using Sentiment Analysis

Shivani Purwar[1], Shweta Verma[2], Smriti Srivastava[3], Ms Deepika Tyagi[4]
Department of Computer Science & Engineering JSS Academy of Technical Education Noida, India

Abstract: This research discusses the ever-changing nature of the stock market by using models such as LSTM and ARIMA to forecast stock prices. It recognizes the volatility and unpredictability of financial markets recognizing the need for robust tools that can accurately predict. Sentiment analysis is used for enhancing the accuracy of stock price prediction, drawing information from various sources like news resources and social media platforms that reflect public sentiments and opinions. We aim to analyze a thorough aspect of factors influencing stock prices by combining these sentiments with the models. The LSTM model is used to study long-term dependencies in stock price trends and ARIMA provides information into the time series components. Combining sentiment analysis helps us to scan the emotional tone around a particular stock and contributes valuable data to the prediction process. The integration between the machine learning models and sentiment analysis offers an extensive approach to predicting stock prices considering both past trends and current public sentiment. This study adds to the work in progress to develop more accurate and adaptive tools for exploring the difficulties of the stock market.
Keywords: LSTM, ARIMA, Sentiment Analysis, Stock Prediction

I. INTRODUCTION

In the intricate landscape of financial markets, predicting stock prices accurately has long been a challenge that intrigues economists, investors, and researchers alike. In the relentless ebb and flow of market dynamics, where vast amounts of data and real-time news influence stock prices [3], gaining a competitive edge becomes essential for success. Today, we embark on a journey to explore a powerful tool that has been revolutionizing the approach to stock market predictions – Sentiment Analysis.
Sentiment Analysis is a methodology that leverages the power of natural language processing and machine learning to analyse public sentiment and emotions expressed in various textual data sources, including news articles, social media, and more [3,6]. By deciphering the collective sentiment of market participants, we can extract valuable insights that aid in making informed investment decisions. The stock market, known for its constant changes and challenges, requires sophisticated tools to navigate its intricate patterns. In our research, we focus on employing two such tools – LSTM and ARIMA models. These tools serve distinct purposes in understanding the complex patterns within stock prices. LSTM, a class of RNN [11,12], specializes in understanding and making sense of relationships and patterns in data. This makes it particularly well-suited for unravelling intricate and convoluted trends that may extend beyond the scope of traditional models.
On the other hand, ARIMA, a classical time series forecasting method, excels in handling historical data and capturing short-term patterns. The synergy between LSTM and ARIMA in our research creates a robust framework for comprehending the multifaceted dynamics of stock prices. While LSTM provides insights into broader trends, ARIMA contributes by capturing short-term fluctuations, offering a comprehensive understanding of the complexities inherent in the stock market. However, the stock market is not solely about numerical indicators and historical data; the sentiments and opinions of individuals equally influence it. This is where sentiment analysis, especially sourced from platforms like Twitter, becomes invaluable. By analysing what people are saying about specific stocks, we aim to capture the emotional aspect of market behaviour. The sentiment analysis process involves collecting and analysing tweets related to specific stocks, using NLP [7] techniques to classify sentiments into positive, negative, or neutral categories. Twitter, as a real-time platform for expressing opinions and reactions, provides a unique window into the collective mood of market participants. The integration of sentiment analysis into our predictive model adds a real-time and qualitative dimension to our understanding of market dynamics. It serves as a pulse check, capturing the emotional undercurrents that may not be immediately apparent through numerical analysis alone. For example, a surge in positive sentiments on social media about a specific stock could precede an uptick in its market performance, providing an early indicator that complements traditional quantitative analysis.
Moreover, our research doesn't stop at the integration of sentiment analysis; it embraces the adaptability of machine learning. The predictive model continuously learns from new data, refining its understanding of market conditions. This adaptability is crucial in a landscape where external factors, from economic shifts to geopolitical events, can swiftly impact stock prices. This adaptability is crucial in a landscape where external factors, from economic shifts to geopolitical events, can swiftly impact stock prices.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

In essence, our research represents a convergence of cutting-edge technologies and traditional wisdom, a fusion of numerical analysis and qualitative insights. The combination of LSTM and ARIMA models addresses the intricate patterns in stock prices, while sentiment analysis from social media provides a real-time pulse of market sentiments. The integration of these components into a cohesive predictive model is a testament to our commitment to developing tools that not only embrace the complexity of the stock market but also provide actionable insights for traders and investors.

Our research is not merely an academic pursuit; it is a proactive response to the challenges posed by the ever-evolving nature of financial markets. The goal is to empower market participants with a tool that not only navigates the complexities of historical data but also taps into the real-time sentiments and perceptions that shape market dynamics. It is a journey towards creating a predictive model that is not confined by traditional boundaries but is agile enough to adapt to the fluid nature of the stock market. In conclusion, our research signifies a pioneering effort to merge cutting-edge technologies with traditional wisdom, forging a predictive model that navigates the multifaceted nature of this market. ARIMA and LSTM models with sentiment analysis from social media platforms represent a dynamic approach. By incorporating real-time sentiments and continuously adapting to evolving market conditions through machine learning, our model strives to empower traders and investors with timely, nuanced insights. It is a commitment to providing a comprehensive tool that not only anticipates market trends but also interprets the intricate interplay of quantitative and qualitative factors shaping the financial landscape.

II. RELATED WORKS

S.No.	Author	Year	Technique Used	Result
1	Sonali Antad et al	2023	Linear Regression	Using historical data and establishing a linear relationship, stock prices were predicted.
2	Jagruti Hota, Bijay K. Paikaray	2022	ANN, SVM, Random Forest	Several different approaches were examined and Random Forest gave the best result.
3	Junaid Maqbool, Ajay Mittal	2023	Traditional Machine Learning Models	MLP regressor was evaluated with diverse sentiments.
4	Mehar Vijh, Arun Kumar, Deeksha Chandola	2020	Artificial Neural Network, Random Forest	The closing price of shares of five different companies was predicted.
5	Payal Soni, Yogya Tewari	2022	Deep Learning algorithms, Neural Networks	Deep Learning algorithms, Neural Networks
6	Shilpa Gite et al	2020	Long Short-Term Memory	CNN, LSTM models were used to make predictions of stock prices using news headlines dataset.
7	Venkata Sasank Pagolu et al	2016	Natural Language Processing	Rise or fall in costs of stocks of company based on the emotions of people was predicted.
8	Xuan Ji, Jiachen Wang and Zhijuin Yan	2021	Text Mining, Deep Learning	Evaluating how the LSTM model behaves in the presence or absence of textual data It improved with text features.

9	Nusrat Rouf, Sparsh Sharma	2022	Generic review, Support Vector Machines	Discoveries from the year 2011 to 2021 were thoroughly examined to predict stock prices.
10	Tinku Singh, Satakshi, Riya Kalra, Suryanshi Mishra	2021	Incremental Learning, technical indicator, deep learning	Using Google Collaboratory real time stock market prices are predicted.
11	Alex Sherstinksy	2020	RNN, LSTM	RNN and LSTM fundamentals are discussed and RNN formulation is derived.
12	Adil Moghar and Mhamed Hamiche	2020	LNN, RSTM	Future stock market prices were predicted using RNN, especially LSTM

III. METHODOLOGY

1) *LSTM-* Its creation was specifically driven by the need to overcome the vanishing gradient problem inherent in conventional RNNs, which impedes their ability to grasp long-term dependencies in sequential data effectively. LSTMs distinguish themselves by their distinctive ability to uphold and update a cell state, enabling the capturing of prolonged dependencies in data. This is facilitated by specialized gates governing the information flow within the network. The crucial components of an LSTM cell encompass the cell state (c), which functions as the cell's memory, and the hidden state (h), which serves as the output for making predictions based on the sequence of inputs.
   The input gate is fundamental for regulating information intake into the cell state. It evaluates the current input and previous hidden state, assigning weights through a sigmoid function. The input gate manages the inflow of information into the cell state, the forget gate decides which information is to be retained or discarded from the cell state, and the output gate oversees the information used to compute the hidden state. Leveraging sigmoid and tanh activation functions within these gates empowers LSTMs to selectively update and output information. This adaptability positions LSTMs as a standard choice for a myriad of tasks involving sequential data, owing to their capacity to effectively model intricate relationships and dependencies in sequential data.
2) *ARIMA-* ARIMA is widely used for predicting trends in sequences of numbers, like those found in finance, economics, and environmental science.

Let's break down ARIMA:

- *Autoregressive (AR):* This part looks at how the current number is related to its past values. It considers how much the current value depends on its previous ones, and the "p" number tells us how many past values to look at.
- *Integrated (I):* This involves making the sequence of numbers more straightforward to work with. The "d" number tells us how many times we need to adjust the dataset for making it easier to understand.
- *Moving Average (MA):* It looks at the relationship between the current number and a calculated error from past values. The "q" number tells us how many past errors to consider.

The standard way we write an ARIMA model is ARIMA (p, d, q), where "p" is about the past values, "d" is about making the data simpler, and "q" is about past errors. ARIMA is good for predicting trends in the short to medium term, especially when the data has a regular pattern. But if the data has long-term trends or complicated patterns, other methods like Seasonal-Trend decomposition using LOESS (STL) or machine learning might work better. Using ARIMA involves picking the right "p," "d," and "q" values based on a careful look at the data, fitting the model to the data, and checking how well it predicts.

IV. CHALLENGES

Based on the provided research paper, some potential challenges or limitations include:

1) *Volatility and Unpredictability of Financial Markets:* Financial markets are inherently volatile and unpredictable. Despite the use of advanced models like LSTM and ARIMA, predicting stock prices accurately remains a challenge due to the dynamic nature of market conditions.

2) *Reliance on Historical Data:* Both LSTM and ARIMA models heavily rely on historical data to make predictions. If market conditions significantly deviate from historical patterns, the models may struggle to provide accurate forecasts.
3) *Noise in Social Media Data:* Social media data, especially from platforms like Twitter, can be noisy and subject to manipulation. Distinguishing genuine sentiments from noise or intentional misinformation poses a challenge.
4) *Generalization of Sentiment Analysis Results:* Sentiment analysis may not universally apply to all market participants. Different groups of investors may interpret information differently, and sentiment analysis might not capture all relevant perspectives accurately.
5) *Model Complexity and Interpretability:* LSTM is a complex neural network architecture, and understanding its internal workings may be challenging. Balancing model complexity with interpretability is crucial for practical application and user acceptance.
6) *Adaptability to Changing Market Dynamics:* While the paper mentions the adaptability of the predictive model to changing market conditions, the specific mechanisms and challenges associated with continuous learning and adaptation are not detailed.
7) *External Factors:* The research acknowledges external factors such as economic shifts and geopolitical events impacting stock prices. However, addressing the challenges of incorporating and accurately interpreting such external factors is a complex task.
8) *Evaluation Metrics:* The paper does not explicitly mention the evaluation metrics used to assess the performance of the predictive models. Choosing appropriate metrics and demonstrating the effectiveness of the proposed approach is essential for validating the research findings.
9) *Ethical Considerations:* The research involves analysing sentiments from social media, which raises ethical considerations regarding user privacy and the responsible use of public opinions.

V. CONCLUSION

In summary, combining sentiment analysis with models like LSTM and ARIMA has shown some promising improvements in predicting stock prices. Each model has its strengths, and when you put together insights from sentiment analysis with time series analysis, it gives a well-rounded approach. But, it's important to recognize that predicting financial markets is tricky, and there are uncertainties. As researchers keep working on this, making these models better and using the latest sentiment analysis methods will be crucial for making stock price predictions more accurate and dependable.

VI. ACKNOWLEDGEMENT

We would like to convey our heartfelt appreciation to Dr Kakoli Banerjee, Head of the Department, for her invaluable guidance and support throughout the preparation of this review paper on stock price prediction using sentiment analysis. Additionally, I extend my heartfelt thanks to my mentor, Ms. Deepika Tyagi, for her continuous encouragement and insightful suggestions that greatly enriched the content of this paper. Their expertise has been instrumental in shaping this work.

REFERENCES

[1] Sonali Antad, Saloni Khandelwal, Anushka Khandelwal, Rohan Khandare, Prathmesh Khandave, Dhawal Khangar and Raj Khanke, “Stock Price Prediction Website Using Linear Regression”, ITM Web of Conferences 56, 05016, 2023.
[2] Jagruti Hota, Sujata Chakravarty, Bijay K. Paikaray and Harshvardhan Bhoyar, “Stock Market Prediction using Machine Learning Techniques”, CEUR Workshop Proceedings, 2022.
[3] Junaid Maqbool, Preeti Aggarwal, Ravreet Kaur, Ajay Mittal and Ishfaq Ali Ganai, “Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach”, Procedia Computer Science 218 (2023) 1067–1078.
[4] Mehar Vijh, Deeksha Chandola, Vinay Anand Tikkiwal and ArunKumar, “Stock Closing Price Prediction using Machine Learning Techniques”, Procedia Computer Science 167 (2020) 599-606.
[5] Payal Soni, Yogya Tewari and Prof. Deepa Krishnan, “Machine Learning Approaches in Stock Price Prediction: A Systematic Review”, Journal of Physics: Conference Series, 2161 (2022) 012065.
[6] Shilpa Gite, Hrituja Khatavkar, Ketan Kotecha, Shilpi Srivastava, Priyam Maheshwari and Neerav Pandey, “Explainable stock prices prediction from financial news articles using sentiment analysis”, DOI10.7717/peerj-cs.340, 2020.
[7] Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda and Babita Majhi, “Sentiment Analysis of Twitter Data for Predicting Stock Market Movements”, arXiv:1610.09225v1 2016.
[8] Xuan Ji, Jiachen Wang and Zhijun Yan, “A stock price prediction method based on deep learning technology”, International Journal of Crowd Science Vol.5No.1,2021 pp. 55-72 Emerald Publishing Limited 2398-7294 DOI 10.1108/IJCS-05-2020-001.
[9] Nusrat Rouf, Majid Bashir Malik, Tasleem Arif, Sparsh Sharma, Saurabh Singh, Satyabrata Aich, and Hee-Cheol Kim, “Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions”, Electronics 2021, 10, 2717.
[10] Tinku Singh, Riya Kalra, Suryanshi Mishra, Satakshi and Manish Kumar, “An efficient real-time stock prediction exploiting incremental learning and deep learning”, Evolving Systems 14:919–937, 2022.

[11] Alex Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network”, Elsevier journal Physica D: Nonlinear Phenomena ”, Volume 404, March 2020.
[12] Adil MOGHAR, Mhamed HAMICHE, “Stock Market Prediction Using LSTM Recurrent Neural Network”, Science Direct Procedia Computer Science 170 (2020) 1168–1173, 2020.