Dissecting the Hype: A Study of WallStreetBets' Sentiment and Network Correlation on Financial Markets

Kevin Wang, Bill Wong, Mohammad Ali Khoshkholghi, Purav Shah, Ranesh Naha, Aniket Mahanti and Jong-Kyou Kim

Abstract The emergence of online investment communities like WallStreetBets (WSB) on Reddit has revolutionised retail trading, characterised by collaborative, meme-infused market influence. This study analyses WSB's ecosystem, examining how social sentiment, network structure, and user interactions impact stock volatility. Our analysis leverages sentiment analysis and network theory on millions of posts to understand community dynamics and their effectiveness in predicting stock prices. This research highlights the rising influence of social media in financial markets, especially with the recent surge in retail investors and their market impact.

Kevin Wang
University of Auckland, Auckland, New Zealand, e-mail: kwan772@aucklanduni.ac.nz

Bill Wong
University of Auckland, Auckland, New Zealand, e-mail: bwon783@aucklanduni.ac.nz

Mohammad Ali Khoshkholghi
Middlesex University, London, United Kingdom, e-mail: a.khoshkholghi@mdx.ac.uk

Purav Shah
Middlesex University, London, United Kingdom, e-mail: p.shah@mdx.ac.uk

Ranesh Naha
Federation University Australia, Melbourne, Australia, e-mail: n.naha@federation.edu.au

Aniket Mahanti
University of Auckland, Auckland, New Zealand, e-mail: a.mahanti@auckland.ac.nz

Jong-Kyou Kim
University of New Brunswick New Brunswick, Canada, e-mail: jongkyou.kim@unb.ca

1 Introduction

In recent years, the convergence of social media and financial markets has birthed a formidable force, one capable of catalysing seismic shifts within the economic landscape. The subreddit WSB, a community within the Reddit platform, exemplifies the potency of this convergence. In January 2021, a short squeeze of GameStop's stock was driven by the subreddit WSB. This collective of retail investors demonstrated the significant influence social media can exert on financial markets. GameStop's stock, heavily shorted at approximately 140% of its public float, experienced an unprecedented rise from $17.25 to over $500 pre-market value. The stock's volatility persisted post-peak, with remarkable surges and declines, exemplifying the impact of digitally coordinated investment strategies on market dynamics.

Our study delves into the narrative and data-driven aspect of the WSB, particularly highlighting events such as the GameStop short squeeze, providing insights to media outlets, financial institutions, and regulatory bodies alike. This phenomenon catalysed a broader discourse about the power of social platforms in shaping market behaviour, leading to concern and calls for regulation. The events also sparked rigorous academic interest in understanding the mechanisms behind such collective actions and their influence on finance systems. The emergence of this 'meme stock' movement, where traders/investors rally around certain stocks promoted on social media, has raised profound questions about conventional financial theories, market infrastructure, and the role of public discourse in the valuation of securities. This study provides a holistic view of how meme stock behaves differently than non-meme stocks by delving into the network structure of the WSB communities, revealing its increased risk of market manipulation.

Our study presents a multi-faceted analysis encompassing sentiment and network analyses derived from a vast collection of Reddit posts and comments. Our research narrows its focus on the WSB subreddit, considering its influence on the stock prices of companies like GameStop (GME) and Tesla (TSLA). By scrutinising the temporal evolution of posting behaviour, user influence, and sentiment trends, we draw correlations that contribute to a nuanced understanding of how social media discourse may reflect or even drive investor sentiment and stock market dynamics. Additionally, we explore the user interactions manifested in the WSB community, revealing the digital Pareto Principle at play and its potential implications for market manipulation. Our study integrates sentiment and network insights from the proposed analysis into Machine Learning (ML) models to investigate their effects on stock price predictions. Our findings highlighted the predictive power provided by both sentiment and network data. It also sheds light on the possible chaos introduced by meme stock sentiments and network structures.

2 Related Work

2.1 Social Media Characterisation

2.1.1 Wealth and Influence Dynamics in Social Media: Disparities, Algorithmic Biases, and Influence Cascades

Guidi et al. [8] examined Blockchain Online Social Media (BOSM), uncovering a highly polarised environment where wealth and influence are concentrated, highlighting the critical need to address socio-economic disparities in such platforms. Glenski and Weninger's study [7] on Reddit revealed how algorithmic biases influence user behaviour and content propagation, indicating the potential of these insights in driving strategic user engagement during important financial events. Bakshy et al. [2] investigated Twitter's role in information cascades, showing the complexity of influencer dynamics in spreading information and suggesting the need for a diversified approach in leveraging online influence, especially in influencer marketing and during periods of market volatility. These studies collectively shed light on the impact of social media on socio-economic aspects and user engagement behaviours.

In a particularly relevant exploration, Thukral et al. [20] studied the user behaviours and interaction patterns on Reddit, unveiling a series of behavioural phenomena and interaction dynamics. The researchers delved deep into the underlying behavioural patterns and dynamics that shape these interactions. The concept of "Limelight Score" is introduced to quantify the depth of discussion around a single comment, revealing that a significant portion of discussions can be centred around one pivotal comment. Another notable introduction is the "Interaction Score," assessing author influence and showing reciprocal behaviour as a key to getting attention on the platform. The findings not only cast light on the varied and multifaceted interactions on Reddit but also pave the way for potential implications in content and advertisement targeting, furnishing a crucial understanding of online social dynamics. Especially in financial contexts, where user discussions can significantly impact market dynamics.

2.2 Social Media and the Financial Market

2.2.1 Social Media's Influence in Financial Markets: Sentiment Analysis and Predictive Insights

Studies reveal the profound impact of social media on financial markets. Analysis of Reddit's r/personalfinance subreddit [21] shows how financial discussions shape user behaviour, with network analysis and topic extraction methods providing insights into collective financial decision-making. Twitter's influence on stock price predictions is illuminated in studies by Jessica and Raymond Sunardi Oetama [14, 19], suggesting the integration of social media sentiment with traditional financial forecasting. Further, research [1] on multi-modal sentiment analysis, including financial video news, indicates the potential for a more comprehensive approach to predicting market fluctuations. The integration of sentiment analysis with algorithmic predictions in market forecasting [17, 11, 9] demonstrates enhanced accuracy in trend prediction. A notable study [18] focuses on Twitter sentiment as a predictive factor for stock market movements, using sentiment analysis and machine learning to correlate public sentiment with stock price changes. These studies collectively emphasise the evolving role of social media in financial market analysis and prediction.

2.2.2 Utilising Reddit Insights for Precise Stock Price Prediction

A highly relevant study [10] ventures into the potential of utilising discussions from WSB to predict stock prices, concentrating on the enhancement of prediction accuracy using techniques of trust filter and a sliding window mechanism. The researchers not only collected sentiment of WSB posts through an individualised approach but also introduced a trust score per record and user, to identify and harness reliable stock-related discussions from the seemingly chaotic pool of user-generated content. In addition to exploring the effects of applying a sliding window approach to aggregate Reddit data, the research acknowledges the prevailing impact of sentiment within financial forums. Employing deep learning research within Natural Language Processing, it adeptly evaluates trends and establishes connections between historical sentiment and financial data. This research not only assists retail investors in navigating and focusing on emerging stock opportunities by capitalising on the collective understanding of credible advisors but also offers a method to determine the optimal window of historical social media information to be considered for adept stock analysis.

2.3 Comparison to prior studies

In contextual comparison with prior works, our study offers several distinct contributions. While previous research has provided valuable insights into the relationship between social media sentiment and financial market predictions, our investigation digs deeper into the network dynamics of WSB. Unlike prior studies that focus predominantly on sentiment analysis, our research integrates network information, such as user influence and interaction patterns, to fortify the predictive models. This comparative analysis not only positions our study within the existing literature landscape but also accentuates our innovative approach.

Our study reveals novel metrics for understanding the gravitas of network influence within WSB, which, to our knowledge, has not been extensively explored before. It is this synthesis of sentiment analysis and network dynamics that delineates our contribution. We are bridging the gap by correlating the quantifiable measures of user influence with stock price movements, which provides a more granular understanding of the market sentiments as reflected in social media interactions. This work, therefore, not only into the unexplored domain of WSB but also proposes a methodologically robust framework that enhances the accuracy and reliability of stock market predictions. The comprehensive insights derived from our research underscore the role WSB plays in relation to the financial markets.

3 Methodology

This section explains the structured approach employed to examine the correlation between social media sentiment, particularly emanating from the WSB forum, and the price movements of stocks such as GME, TSLA, and S&P 500 (SPY). It explicates the phases of data collection, sentiment analysis, network analysis, and predictive modelling, along with their respective evaluation metrics.

3.1 Data Collection

The workflow of data collection involved using the Python Reddit API Wrapper (PRAW), Pushshift API, and The-eye Reddit archive to collect historical posts and comments from the subreddit WSB (Fig. 1a). From this endeavour, around two million posts were collected from WSB. Another two million comments pertinent to TSLA and GME were extracted from The-eye archive, which hosts an exhaustive repository of WSB comments preceding March 2023. The extraction of the two million comments is based on the inclusion of words “$TSLA”, “$GME”, “Tesla” or “Gamestop” in the body of each comment. Hence any comments mentioning these words will be extracted from the historical WSB comment dataset. This extraction enables a more targeted analysis of stock-specific discussions. To further refine the dataset for subsequent network analysis, we leveraged the relational structure of Reddit’s data to fetch the authors of parent posts for each comment. By correlating each comment with its parent post ID, we could accurately attribute comments to their original discussions and extract the associated parent post author, a critical step for the construction of the communication graph detailed later in the methodology.

Financial datasets encompassing daily trading metrics for the stocks TSLA, GME, and SPY were obtained from the Yahoo Finance API, complementing the social media data. Data cleansing and validation comprised the elimination of duplicate records and the dropping of Not-a-number (Nan) records, which together ensured the robustness of the data foundation for the study.

3.2 Characterisation Metrics

The sentiment of each post and comment was quantified using FinBERT[6], which exhibited superior accuracy (74%) over VADER [3] (50%) in our preliminary tests on a financial dataset [15]. The Louvain algorithm was applied for community detection within the WSB network [5], and the Google PageRank algorithm assessed the relative importance of each user within the WSB community [16]. Temporal network changes were tracked via metrics such as the number of edges, nodes, and average degrees, providing a dynamic view of community interaction over time.

The study also involves building ML models for stock price predictions leveraging financial attributes, sentiment data and network data. The financial attributes include open, high, low, and close prices and volume of a given stock. The sentiment data is the daily net sentiment score of the given stock in WSB. And the network data is the PageRank importance score of authors in WSB. The models used are Random Forest (RF) [4], Multilayer Perceptron (MLP) [13] and Long-short Term Memory (LSTM) [12]. These models are evaluated and compared using the Root-mean-square Error (RMSE) metric.

3.3 Data Overview

The data compiled presented a robust foundation for analysis, summarised in the following table:

Table 1 WSB data overview

Dataset	Number of Posts	Number of Comments	Number of Authors	Average comments per post
WSB (TSLA)	42528	72766	30554	2.4
WSB (GME)	226330	2000847	73974	27.0
WSB (all)	2275310	96317	634201	29.0

The collected WSB dataset is uploaded to Kaggle[1]. Table 1 summarizes data from the WallStreetBets (WSB), covering three distinct datasets: WSB (TSLA), WSB (GME), and WSB (all). The WSB (TSLA) dataset includes 42,528 posts, 72,766 comments from 30,554 authors, averaging 2.4 comments per post. In contrast, the WSB (GME) dataset is larger, with 226,330 posts, 2,000,847 comments from 73,974 authors, and an average of 27.0 comments per post. The most comprehensive dataset, WSB (all), encompasses 2,275,310 posts, 96,317 comments from 634,201 authors, averaging 29.0 comments per post.

[1] Kaggle WSB Dataset Link: `https://www.kaggle.com/datasets/kevinwang313/wallstreetbets-dataset`

4 Results

4.1 *Sentiment and Network Insights Mostly Enhance Model Performance*

The outstanding pattern recognition ability of ML models led to their domination in stock price/trend prediction. It certainly is a big interest for many researchers and traders to investigate methods and techniques that can further enhance ML model performances due to the financial benefits it could potentially bring. We conducted model evaluations on RF, FNN, and LSTM to assess the impact of sentiment data and network data on improving model performance, we started with a base model that solely utilises financial attributes. Subsequently, we integrated sentiment scores and PageRank importance scores into the model input for further evaluation. The specific configurations of these models are detailed as follows:

- RF exclusively with Financial Attributes (RF-FA)
- RF integrating Financial Attributes and Sentiment Score (RF-FA-SS)
- RF amalgamating Financial Attributes with a PageRank-weighted Sentiment Score (RF-FA-SS-PW)
- FNN focusing solely on Financial Attributes (FNN-FA)
- FNN combining Financial Attributes and Sentiment Score (FNN-FA-SS)
- FNN merging Financial Attributes with a PageRank-weighted Sentiment Score (FNN-FA-SS-PW)
- LSTM centred on Financial Attributes (LSTM-FA)
- LSTM integrating Financial Attributes and Sentiment Score (LSTM-FA-SS)
- LSTM synthesising Financial Attributes with a PageRank-weighted Sentiment Score (LSTM-FA-SS-PW)

Table 2 RSME of all models for predicting TSLA and GME stock price.

	TSLA	GME		TSLA	GME
RF-FA	13.90	2.22	FNN-FA-SS-PW	10.41	2.37
RF-FA-SS	13.53	2.08	LSTM-FA	11.28	1.98
RF-FA-SS-PW	13.01	2.07	LSTM-FA-SS	10.51	1.67
FNN-FA	11.46	1.96	LSTM-FA-SS-PW	10.39	1.69
FNN-FA-SS	13.33	1.65			

A clear observation from our experiments was the enhancement in model performance upon the inclusion of sentiment data. This is supported by the uniform decline in RMSE values when comparing models equipped solely with Financial Attributes against those equipped with both Financial Attributes and Sentiment Scores. Such consistent improvements imply a correlation between WSB sentiment and stock price dynamics. While our sentiment analysis indicated a lagging sentiment effect on stock price, it showcased minimal evidence of sentiment data leading stock price movements. This prompts us to postulate the existence of a non-linear association between sentiment data and stock price movement, a relationship potentially not captured by Pearson correlation detection. A fascinating discovery from the model evaluation pertains to the distinct impacts of PageRank Weighting (PW) on TSLA and GME predictive modelling. For TSLA, the incorporation of PW persistently mitigated the RSME across all models with a 9% decrease in RMSE on average. In contrast, all GME models had no statistically significant enhancements by incorporating PW. This discrepancy shows the possible distinguished user behaviours between meme stocks and traditional stocks, inferring a notable variance in network interactions within WSB. The chaotic dynamics of meme stock networks on WSB may lead to more volatile and potentially deceptive sentiment data, introducing additional noise and thereby obstructing model generalisation. Conversely, the sentiment distribution for traditional stocks like TSLA appears more stable and resilient to manipulation, rendering the insights from influential voices invaluable. Our results resonate with multiple studies affirming the efficacy of sentiment data in enhancing model performance [1, 17, 11, 18, 9, 10]. While WSB sentiment may not emerge as a singular influential factor upon which retail traders should depend, its incorporation increased the robustness of price-prediction models. Our study delineates itself by unravelling the utility of PW in refining sentiment data for predicting conventional stock prices. Employing PW amplifies the significance of individuals possessing heightened influence and followership. In typical stock discussions, these influential individuals often demonstrate a profound comprehension of stocks and predominantly disseminate quality insights, thereby enhancing the calibre of the sentiment data. However, in the context of meme stock discussions, these influencers occasionally resort to exaggeration to instigate excitement, thereby infusing noise into the sentiment data. Huynh et al. utilised trust filters on WSB data, revealing notable improvements in their findings [10]. Trust filters were employed to exclude data originating from bots and less active users by scrutinising their posts and comments. While this approach shares some similarities with our approach, our methodology transcends the trust filter technique by integrating network information. Rather than discarding data, we retain all sentiment data but modulate its significance based on a network analysis of all individuals discussing a specific stock. Our outcomes indicate that utilising PageRank scores can be particularly advantageous for conventional stocks like TSLA. The robust performance of our LSTM-FA-SS-PW model indicates its potential utility for both retail and institutional traders in enhancing predictions. Furthermore, we postulate that incorporating SS and PW into LSTM models can enhance other systems, such as those employing reinforcement learning (RL) for stock trading. For instance, numerous RL systems leverage LSTM models to predict stock prices. The addition of SS and PW to these RL systems could potentially enhance their performance. Our findings suggest that using PageRank weighted sentiment could enhance model performance for TSLA price prediction. It is important to note that we suggest that there are potential enhancements to model performances by using PageRank weighted sentiment. However, due to the scope of this research, this technique needs to be tested on more stocks to prove its consistency and robustness.

4.2 User Interactions Exhibit Notable Disparities between

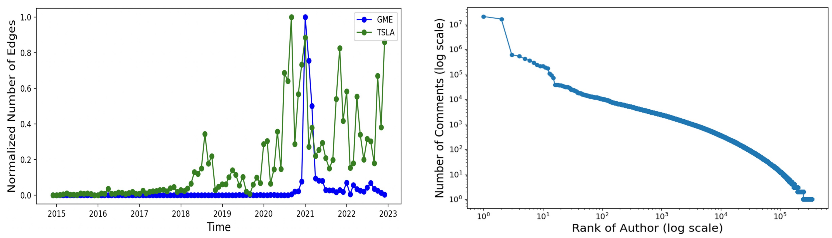


Fig. 1 (a) Normalised number of edges over time for TSLA and GME (b) Zipf's distribution showing authors and number of comments

The temporal analysis reveals interesting insights into the WSB networks. It shows how certain metrics change over time, therefore showing the growth and progress of the network which could reveal interesting patterns of the network. Fig. 1a shows a comparison between the edge number of GME and TSLA network graphs over time. The number of edges for both the GME network and TSLA network are normalised to highlight the difference in its patterns. An edge in the network graph represents a comment which is the primary method of interaction in Reddit. For stocks like GME, these graphs illuminate a significant surge in activity during events like the short squeeze, which then quickly returns to pre short squeeze level. In contrast, stocks such as Tesla demonstrate a steadier and more consistent growth in interactions. Our data shows that meme stocks like GME, are predominantly initiated by a limited number of authors. In comparison, interactions around more conventional stocks are broader and more evenly distributed. There is a noticeable difference between user behaviours of meme stocks and traditional stocks in WSB. Often, individuals or groups will artificially inflate the value of these stocks in what is known as "pump and dump" tactics. They will generate hype around a stock, driving up its price, only to sell their shares once the price peaks, leaving other investors at a loss. When we see an unusually high level of activity or discussion around a particular stock, it could be a warning sign. It might indicate that the stock is not fairly evaluated, and there is a risk that its value could plummet soon after. This is captured by Fig. 1a, where the GME interactions had an unusual surge, and it dropped back to normal levels soon after. Retail investors must be aware of these patterns and exercise caution when considering investments in such stocks. This analysis shows the distinct user behaviour of meme stocks and non-meme stocks. Highlighting the potential market manipulations. However, on the other hand, the GME short squeeze does not represent the characteristics of all meme stocks. A more comprehensive study on a wide range of meme stock behaviours would be beneficial in crafting regulations and defence mechanisms for preventing market manipulations.

4.3 *The WSB Community Demonstrates Pronounced Susceptibility to Manipulation through Influential Dynamics*

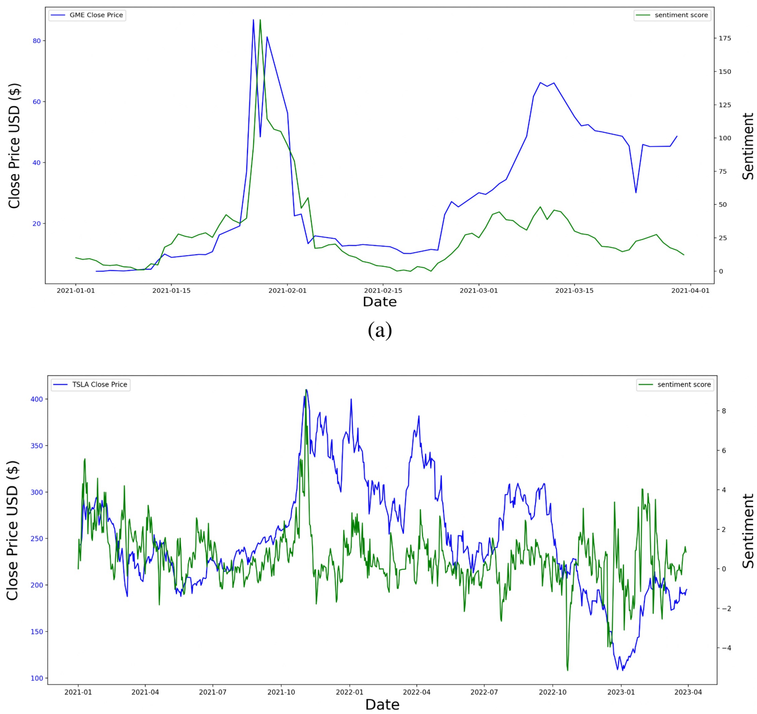
To verify the constant Pareto Principal findings in various studies [2] and to identify dominant actors within the WSB community, network graphical analyses were employed. In this context, the Pareto Principle or the 80-20 rule, oftentimes observed within various systems, is explored alongside the PageRank algorithm to navigate through the intricate web of user interactions and influence within WSB. Our results showed that WSB not only adhered to the Pareto Principle but exhibited an even more skewed interaction distribution, with a mere 20% of user posts captivating a staggering 97.70% of comments, far surpassing the conventional 80-20 distribution. This phenomenon further aligns with Zipf's distribution (Fig. 1b), characterised by a slope of -1.72, signifying an extreme imbalance in the distribution of user interactions and influence. Such an extreme manifestation of the Pareto Principle in WSB implies the susceptibility to market manipulation by a small set of top authors wielding considerable influence over the majority of its user base. The execution of PageRank analysis highlighted that a subset of authors, specifically "bigbear0083" and "OPINION_IS_UNPOPULAR", exert significant influence within the WSB community. These highly influential figures introduce a nuanced complexity and potential for bias. The capacity for influential entities, especially moderators, to amplify or perhaps suppress particular narratives holds the potential to significantly skew the WSB network's dialogues, thereby introducing a heavy bias within the forum. Our results indicate a high potential for manipulation; however, it's crucial to recognise that numerous influential individuals may offer unbiased, truthful information, including valuable financial advice and insightful market analysis. These influencers might possess a strong sense of integrity that deters them from engaging in manipulation. Ideally, in a finance forum, this would be the norm. Nevertheless, occurrences such as the GME short squeeze and the emergence of various meme stocks particularly increased the risk of market manipulation within the WSB community.

4.4 Sentiment Pattern Shows Moderate Correlation and Lagging Effect Compared to Price Movements

Our sentiment analysis focused on potential correlations between WSB sentiment data and stock price movements, particularly for enhancing stock price predictions. We analysed daily net sentiment data, summing up sentiment scores from comments related to specific stocks, against daily stock prices. We found a peak Pearson correlation of 0.40 between GME stock price changes and 7-day Exponential Moving Average (EMA) of GME net sentiment (Fig. 2a), during the January 2021 short squeeze. For TSLA, the peak correlation was 0.29, observed between the 7-day EMA of net sentiment and the subsequent day's closing price (Fig. 2b). Both GME and TSLA sentiment indicators showed a lagging trend behind stock price movements, indicating stock prices influenced WSB sentiment rather than vice versa. While recognising the inherently lagging nature of the EMA indicator, even the most up-to-date net sentiment score manifested this lagging tendency. This suggests that shifts in GME's stock price were precursors to the sentiment alterations within WSB posts related to GME, thereby influencing community reactions based on stock performance. The study from [1] also unveiled a lagging correlation between Reddit activity and stock price changes in a one-to-four-day period, lacking any linear relationship between sentiment and stock price changes. The similarity between our results and the study provides more reliability in the validity of the sentiment analysis findings and implications. While the intention was to harness the results of sentiment analysis to enhance stock price prediction, our results show the opposite relationship, where price has better predictive power. Although using WSB sentiment data for stock price prediction may not improve performance, the sentiment analysis predominantly concentrated on linear correlations. Hence, there is no evidence to reject the possibility for the ML models to discover non-linear predictive patterns, which enhances prediction performance.

5 Discussion

The integration of sentiment and network data into machine learning models for stock price prediction, as demonstrated in our study, has clear implications for various stakeholders in the financial markets. The reduced RMSE in models incorporating sentiment data points to the value of trader sentiment, particularly from forums like WSB, in understanding stock dynamics. This is crucial for investors and traders who seek to harness alternative data to gain an edge in the market. For meme stocks like GME, our results indicate that sentiment data may add noise rather than value, potentially due to the stock's volatile nature and the influence of a few key actors. In contrast, the predictive accuracy for more traditional stocks like TSLA improves with the inclusion of sentiment and network insights, highlighting the stability and informational value of the sentiment in these markets. Our study contributes to the



(a)

(b)

Fig. 2 (a) GME Close Price and 7-day EMA of Sentiment (b) TSLA Close Price and 7-day EMA of Sentiment

literature by elucidating the variable impact of sentiment data on different types of stocks. The use of PageRank-weighted sentiment scores is a novel approach that refines the influence of sentiment based on the credibility and network position of the individuals within the WSB community. This could be particularly beneficial for financial analysts and researchers who require a more granular understanding of market sentiment. Our findings could be of significant interest to regulatory authorities. The notable variations in user engagement, particularly with meme stocks, highlight the need for regulatory measures to deter market manipulation. Although much of our analysis points to a centralised character within the WSB community, it is possible that key influencers are offering impartial and honest financial perspectives. However, incidents like the GME short squeeze and the creation of various meme stocks emphasise the urgency for improved regulation to reduce the susceptibility of WSB to market manipulation. In terms of practical applications, our findings can enhance algorithmic trading strategies by integrating sentiment analysis, especially when using advanced ML models like LSTMs. For the research community, our approach can be replicated or expanded upon with other datasets or within different market conditions to further validate and build on our conclusions.

6 Conclusion

In conclusion, this research has shed light on the network structures, user interactions, and sentiment within the WSB community on Reddit. By integrating insights from network analysis and sentiment analysis into machine learning models, the research aimed to improve the prediction of the financial market. The network analysis has shown a heavily skewed distribution of user interactions within WSB, where a small subset of influential individuals exert substantial influence over most users. This phenomenon is a clear representation of the Pareto principle within social media platforms, which was also reported in prior studies [8], thus not only supporting our findings but also contributing to a holistic understanding of user interactions across online social platforms. The sentiment analysis indicates a moderate correlation between WSB sentiment data and stock price movements, with a lagging effect observed. This suggests that shifts in stock prices are precursors to changes in sentiment within the WSB community, degrading the predictive power provided by the sentiment data. The integration of insights from both network and sentiment analyses into machine learning models shows enhancement in the predictability of stock price movement, particularly with traditional stocks like Tesla, thereby validating the efficacy of this integration. However, it also brought to light the nuanced and distinct behaviours and volatile sentiment distribution associated with meme stocks like GME, providing a fresh perspective into WSB user behaviour and their influence on market dynamics. The discovery that the proposed techniques amplify the performance of price prediction is significant as it offers tangible advantages to future researchers by enabling more accurate and effective financial forecasting. Overall, these findings have addressed all research objectives by concluding the above results regarding the WSB network, sentiment, and its implications on stock price prediction. The exploration of this study is limited to the sentiment and networks of GME and TSLA within the WSB community due to time and resource constraints. Future research could broaden the scope of this study and conduct a comprehensive network and sentiment analysis across a wider spectrum of stocks within WSB. This could furnish additional evidence or perhaps offer counterarguments regarding the disparate behaviours of conventional and meme stocks. Additionally, the strategy of using sentiment and network insights into predictive models need not be restricted to price prediction alone. This methodology might be effectively utilised in cutting-edge models, including reinforcement learning for stock trading, potentially enhancing the model's capacity to optimise profits.

References

1. Alzazah, F., Cheng, X., Gao, X.: Predict market fluctuations based on the tsi and the sentiment of financial video news sites via machine learning. In: 2023 15th International Conference on Computer and Automation Engineering (ICCAE), pp. 302–308. IEEE (2023)
2. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone's an influencer: quantifying influence on twitter. In: Proceedings of the fourth ACM international conference on Web

search and data mining, pp. 65–74 (2011)
3. Berri, A.: Sentimental analysis using vader (2020). URL https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664
4. Biau, G., Scornet, E.: A random forest guided tour. Test **25**, 197–227 (2016)
5. De Meo, P., Ferrara, E., Fiumara, G., Provetti, A.: Generalized louvain method for community detection in large networks. In: 2011 11th international conference on intelligent systems design and applications, pp. 88–93. IEEE (2011)
6. Genc, Z.: Finbert: Financial sentiment analysis with bert (2020). URL https://medium.com/prosus-ai-tech-blog/finbert-financial-sentiment-analysis-with-bert-b277a3607101
7. Glenski, M., Weninger, T.: Predicting user-interactions on reddit. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp. 609–612 (2017)
8. Guidi, B., Michienzi, A., Ricci, L.: Assessment of wealth distribution in blockchain online social media. IEEE Transactions on Computational Social Systems (2022)
9. Guo, X., Li, J.: A novel twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency. In: 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 472–477. IEEE (2019)
10. Huynh, D., Audet, G., Alabi, N., Tian, Y.: Stock price prediction leveraging reddit: The role of trust filter and sliding window. In: 2021 IEEE International Conference on Big Data (Big Data), pp. 1054–1060. IEEE (2021)
11. Kalra, S., Prasad, J.S.: Efficacy of news sentiment for stock market prediction. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 491–496. IEEE (2019)
12. Memory, L.S.T.: Long short-term memory. Neural computation **9**(8), 1735–1780 (2010)
13. Noriega, L.: Multilayer perceptron tutorial. School of Computing. Staffordshire University **4**(5), 444 (2005)
14. Oetama, R.S., et al.: Sentiment analysis on official news accounts of twitter media in predicting facebook stock. In: 2019 5th International Conference on New Media Studies (CONMEDIA), pp. 74–79. IEEE (2019)
15. Patil, R.: Financial news sentiment analysis using finbert (2023). URL https://github.com/Raviraj2000/Financial-News-Sentiment-Analysis-using-FinBERT
16. Rogers, I.: The google pagerank algorithm and how it works (2002)
17. Sandeep, U., Vardhan, T.V., Yaswanth, B., et al.: Cracking the code: Unleashing the power of sentiment analysis & ml for moroccan stock market forecasting. In: 2023 8th International Conference on Communication and Electronics Systems (ICCES), pp. 1274–1278. IEEE (2023)
18. Sasank Pagolu, V., Nayan Reddy Challa, K., Panda, G., Majhi, B.: Sentiment analysis of twitter data for predicting stock market movements. arXiv e-prints pp. arXiv–1610 (2016)
19. Shah, D., Isah, H., Zulkernine, F.: Predicting the effects of news sentiments on the stock market. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 4705–4708. IEEE (2018)
20. Thukral, S., Meisheri, H., Kataria, T., Agarwal, A., Verma, I., Chatterjee, A., Dey, L.: Analyzing behavioral trends in community driven discussion platforms like reddit. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 662–669. IEEE (2018)
21. Thukral, S., Sangwan, S., Chatterjee, A., Dey, L., Agrawal, A., Chandra, P.K., Mukherjee, A.: Understanding how social discussion platforms like reddit are influencing financial behavior. In: 2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 612–619. IEEE (2022)