IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 August 2024, accepted 17 August 2024, date of publication 26 August 2024, date of current version 10 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3450311

RESEARCH ARTICLE

Investigating Stock Prediction Using LSTM Networks and Sentiment Analysis of Tweets Under High Uncertainty: A Case Study of North American and European Banks

LUCA BACCO[1,2], LORENZO PETROSINO[3], DOMENICO ARGANESE[1], LUCA VOLLERO[1], (Member, IEEE), MARCO PAPI[4], AND MARIO MERONE[3], (Member, IEEE)

[1]Research Unit of Computer Systems and Bioinformatics, Department of Engineering, University Campus Bio-Medico di Roma, 00128 Rome, Italy
[2]ItaliaNLP Laboratory, National Research Council, Istituto di Linguistica Computazionale "Antonio Zampolli," 56124 Pisa, Italy
[3]Research Unit of Intelligent Health Technologies, Department of Engineering, Università Campus Bio-Medico di Roma, 00128 Rome, Italy
[4]Research Unit of Automatic Control, Department of Engineering, Università Campus Bio-Medico di Roma, 00128 Rome, Italy

Corresponding author: Mario Merone (m.merone@unicampus.it)

ABSTRACT The investigation focused on the prediction of stock prices amidst significant macroeconomic and geopolitical volatility, particularly targeting North-American and European banks in 2022 — a year marked by intense economic shocks, including inflation, geopolitical tensions, and supply chain disruptions. A multidimensional approach was employed, integrating advanced Artificial Intelligence (AI) techniques such as Recurrent Neural Networks (RNNs) and sentiment analysis, utilizing a comprehensive dataset that includes traditional financial metrics and sentiment-driven data from social media, specifically Twitter (recently renamed *X*). By employing LSTM and FinBERT models, the study revolved around several key analyses: assessing the impact of different market conditions across the US and EU; exploring the potential benefits of data aggregation from multiple banks within these markets; examining the influence of varying historical data spans on model performance; and integrating sentiment analysis to capture the nuanced influence of public sentiment on stock movements. The findings indicate that market-specific dynamics significantly affect the predictive models, with higher inter-bank correlation observed in the US compared to a more fragmented European market. Additionally, models incorporating recent data streams and public sentiments tend to outperform those relying on traditional, longer historical data.

INDEX TERMS Artificial intelligence, natural language processing, sentiment analysis, stock price forecasting, time series.

I. INTRODUCTION

Recent years have seen unprecedented macroeconomic volatility and geopolitical disruptions that have profoundly affected the banking sector. The traditional stability of this sector was significantly challenged, resulting in banks trading at discounted values relative to other sectors. The impact of these disruptions has been notably severe in 2022, a year marked by robust revenue growth but overshadowed by multiple economic shocks, including inflationary pressures, geopolitical tensions, and supply chain disruptions exacerbated by the ongoing effects of the pandemic.

At least four shocks are affecting banks globally: (i) *Macroeconomic shock*, as the rising of inflation and the likelihood of recession are testing central banks; (ii) *Asset value shock*, including the devaluation of fintech and cryptocurrencies (as the bankruptcy of some high-profile crypto organizations); (iii) *Energy and food supply shock* related to the war in Ukraine, are contributing to inflation and putting millions of livelihoods at risk; (iv) *Supply chain shock*

The associate editor coordinating the review of this manuscript and approving it for publication was Akin Tascikaraoglu.

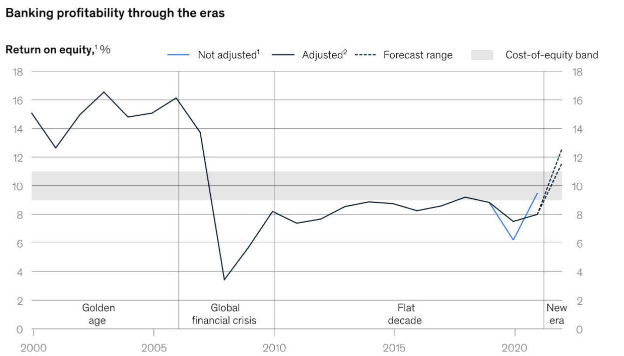

FIGURE 1. Banking profitability through the eras, reported in terms of percentage of return on equity, adjusted from 2000 till 2021, not adjusted from 2019 till 2021, and forecasted range from 2021 till 2022. The figure reports the divisions from the Golden era to the New one.

began during the first pandemic lockdowns continues to affect global markets.

These shocks have revealed and magnified the vulnerabilities of banks, driving a divergence in their performance based on geographic location, funding profiles, and operating models. Despite the return on equity for banks reached a 14-year high in 2022, as shown in Figure 1 with expected return on equity between 11.5 and 12.5 percent,[1] many still reported returns below their cost of equity, highlighting the ongoing challenges within the sector. The result of these pressures represents an increase in the divergence trend between banks observed during the period taken into consideration for the present study, with outcomes varying considerably, depending on funding profile, geography, and operating model.

This scenario of high uncertainty provides a unique backdrop for this study, which aims to explore the interplay of multiple factors influencing bank stock prices. The main research questions of this study are to investigate how advanced AI techniques can enhance the prediction of stock prices under the impact of several factors. To navigate this intricate landscape, this work leverages advanced Artificial Intelligence (AI) techniques, including Recurrent Neural Networks (RNNs), to forecast stock price movements under different circumstances.

The presented comprehensive analysis spans, indeed, across several dimensions: (i) analyzing two major economic market areas — North-American (US) and European (EU) — to account for different geopolitical conditions. (ii) exploring whether aggregating data from multiple banks within these regions can enhance predictive performance. (iii) examining the impact of varying historical data availability on model outcomes during both training and inference phases. (iv) integrating sentiment analysis of tweets from influential CEOs and financial analysts, recognizing that market movements are increasingly driven by public sentiment, which can shift rapidly following newsworthy events or influential commentary.

This approach not only captures the intricate time series patterns of stock prices but also incorporates sentiment analysis to gauge how investor attitudes and emotional responses, often expressed through platforms like Twitter, influence market dynamics. Stock price dynamics remain a critical area of study due to their significant implications for the economy and risk management. While traditional research has focused on financial and technical indicators to assess investor interest and predict stock trends [1], the influence of emotional and mood-driven decisions has become a pivotal element of modern financial analyses [2], [3], [4]. To support the research, an aligned Twitter (recently renamed *X*) dataset was assembled from 2022, featuring posts by prominent CEOs and financial analysts, aiming to enrich the understanding of the complex interplay between sentiment and stock market fluctuations.

The motivation for this research stems from the increasing complexity and interconnection of the global financial markets, which are influenced by a variety of factors including macroeconomic policies, geopolitical events, and public sentiment. Traditional models that rely solely on historical financial data may not fully capture the dynamic nature of these influences. Therefore, there is a need to develop more sophisticated models that can incorporate a broader range of data sources to improve predictive accuracy. This study thus aims to deepen the insight into the complexities of developing intelligent systems for stock price prediction, which would have significant implications for the development of more robust and adaptive predictive models, providing a valuable resource for newcomers venturing into this intricate field.

The paper is structured as follows. In the next section, a review of the existing literature on stock price prediction and financial sentiment analysis is presented, while in Section III the materials and the methods employed in the study are introduced, describing dataset, model architecture, experimental setups, and evaluation metrics. Then, the results and findings are presented and discussed in Section IV, and the key insights and suggestions for future research are resumed in the Conclusion section.

II. RELATED WORK

The theory of random walks in the stock market, originally proposed by Fama [5], [6], suggests that stock price fluctuations can be viewed as random processes. Despite substantial evidence indicating that relying solely on past price trends for future price predictions does not yield consistent profits [7], [8], the prediction of stock price movements remains an active and compelling research area in financial engineering, being a hot topic for the research in this field lately [9], [10].

Current research predominantly focuses on developing models that predict changes or trends in stock prices, often framed as classification tasks [11], [12], rather than attempting to predict exact future prices. A prevalent method in this

[1]Source: S&P Global https://www.spglobal.com/en/

realm is technical analysis, which leverages stock price data as model inputs. This data typically includes daily opening and closing prices sourced from public databases such as *Nasdaq*[2] [13], *Bloomberg*[3] [14], and *Yahoo Finance*[4] [15]. With the evolution of machine learning technologies, particularly deep learning and enhanced computational capabilities, there has been a significant increase in the use of neural network-based models like Convolutional Neural Networks (CNNs) for analyzing stock prices [16], [17].

Among the various types of neural networks used in this context, RNNs like Long Short-Term Memory (LSTM) networks, have been extensively explored for analyzing market behavior and beyond. Their inherent ability to handle sequential data makes them particularly suitable for fields involving temporal data, ranging from healthcare to economics [18]. Specifically, LSTMs have proven successful in developing forecast systems for predicting financial market volatility [19], and have been effectively combined with CNNs [20], traditional econometric models [21], [22], and sentiment analysis techniques [23], [24]. The continued success of these networks, even amidst the emergence of more complex architectures (e.g., Transformers), underscores their suitability for the analysis intended in this paper.

Moreover, the role of textual data in stock price prediction has become increasingly prominent [25], as investors' sentiment and financial textual sentiment are interconnected and influence each other. Traditionally, investor sentiment is gathered through surveys that regularly solicit opinions on market conditions. However, as the availability and quality of digital data grow, researchers have begun to explore how news articles and social media sentiment can impact stock market dynamics. Financial news, for instance, has proven valuable for training predictive models by extracting sentiment scores, which have been shown to improve model performance beyond that achievable with historical price data alone [24], [26], [27]. Platforms like *X* (a.k.a, Twitter), however, have become crucial for sentiment analysis, with several studies demonstrating a correlation between stock price movements and public sentiment expressed about companies online [28], [29], [30].

The sophistication of sentiment analysis techniques has also advanced [31], for either binary (positive vs negative) assessments or more complex emotional dimensions. These technique-driven advances have also been reflected in the finance domain [25]. Indeed, the financial sentiment analysis panorama has rapidly shifted from lexicon approaches to machine and deep learning ones, overcoming the main drawbacks of the former ones, which are usually less accurate that learning-based methods and require significant human efforts for building domain-specific lexicons [32], [33], [34].

[2]https://www.nasdaq.com/

[3]https://www.bloomberg.com/

[4]https://finance.yahoo.com/

This progression has been further propelled by the introduction of the Transformers architecture in Natural Language Processing, exemplified by specialized models such as FinBERT, which was specifically tailored for financial sentiment analysis [35] and was employed in several recent works to extract information from financial texts, like named entities [36], [37]. These application-driven developments highlight the growing potential of integrating textual data with traditional quantitative inputs to enhance the accuracy of stock price predictions.

Our study builds on all these advancements, integrating sentiment analysis as a component of a broader comparative analysis aimed at evaluating the performance of AI models in predicting stock price trends under varying conditions.

III. MATERIALS AND METHODS

A. DATASET

The collected dataset comprehends two subsets reporting either the stock price series over time (*Finance dataset*) or the public tweets of relevant financial accounts (*Twitter dataset*). Both the two subsets refer to a total of 26 banks stocks, under the period span from *December,* 31^{st} *2021* to *December,* 31^{st} *2022*. In particular, to include banks from both the North-American and European markets, the banks were selected among the highest in terms of market capitalization associated with either the *Federal Reserve* (17 banks) or the *European Central Bank* (9 banks). Market capitalization indicates the size and scope of one bank in relation to the others. Generally, banks with higher market capitalization are considered larger and more important than those with lower market capitalization, which reflects on the investors confidence.

Each sample in the built dataset was structured as a $\beta \times n$ matrix, for which its size is determined by the number of β timestamps (i.e., the sequence length) and the number of features n for each timestamp (i.e., the features extracted from the financial time series and, eventually, the features sentiment features extracted from the tweets).

1) FINANCE DATASET

For each selected bank under analysis, the stock information was extracted from *Yahoo Finance*. For each stock, at each time (day) t, features regarding the daily opening and closing prices, the daily trading volume, the highest price, and the lowest price were retrieved. It is worth noting that, being the stock market recorded during the working days (from Monday to Friday, excluding several holidays throughout the year), the time series under analysis do not completely cover the 365 days of the calendar year. Then, given the stock price value $S_t \in \mathbb{R}$, for each time t, S_t was remapped into categorical classes y_t, reflecting either a positive or a negative trend, as follows

$$y_t = \begin{cases} +1, & \text{if } \Delta = S_t - S_{t-1} > 0 \\ -1, & \text{otherwise} \end{cases}$$

2) TWITTER DATASET

To constitute the corpus of financial tweets associated with the selected banks, the process began by identifying the most active profiles in publishing bank-related information through the use of the *TrackMyHashtag* online service.[5] A total of 34 profiles belonging to important financial analysis individuals and organizations were identified, including CEOs and executives.[6] Then, tweets and their related features (i.e., retweets, likes, and dislikes) were collected from those accounts via the *Apify* scraping framework,[7] and then polished. In particular, only informative tweets written in the English language were included, by filtering out tweets either automatically identified as belonging to other languages,[8] consisting only of emoticons, or containing Wordles-like texts.[9] Thus, the collected dataset consists of $\approx 14k$ tweets, for a total weight of 43.1 MB of textual data.

All the collected tweets were published in the same time frame used for the collection of financial data. However, since tweets can be published every day, it was necessary to align the tweets published during weekends (or other holidays) with the ones posted on Mondays (or the day after the holiday ending). After being properly tokenized, each tweet was given in input to *FinBERT* [35], a well-known Large Language Model specifically designed for sentiment analysis in finance,[10] to obtain a triplet of values corresponding to the positive, neutral, and negative sentiments expressed by the text. Then, given the empirical literature showing financial market contagion, i.e., interdependence between financial markets [38], [39], [40], and the relatively high correlations shown by the several banks (Section III-A), the sentiment triplets were averaged across all the tweets posted on a given day by weighting each tweet by the number of its retweets and likes, thus modeling the importance of each tweet by its diffusion across the social platform.

B. MODEL

Building on recent advancements in stock price forecasting [24], [26], [29], the study employs an LSTM network as the foundational architecture of the model. LSTMs represent an evolution of traditional RNNs, distinguished by their capacity to capture and learn dependencies over extended time periods. This is made possible through their sophisticated design, which includes memory cells and gating mechanisms, enabling the network to selectively retain or disregard information throughout lengthy data sequences [41]. With the integration of on-top fully-connected layers, this ability makes them particularly suitable for tackling time-series forecasting and sequence classification tasks.

To optimize the model's performance, a hyper-parameter search was conducted, as detailed in Appendix A. Additionally, to ensure the robustness of the model, an early stopping mechanism was implemented during training to prevent overfitting, halting the training process if there is no decrease in the loss value on the validation set for 200 consecutive epochs. The model was trained from scratch by randomly initializing its weights once for all the experimental setups presented next.

C. EXPERIMENTAL SETUPS

To thoroughly investigate the proposed methodologies, a series of experiments were conducted by varying several factors that could influence the model's performance, and thus assess their impact. To ensure consistency in the experiments and enable a straightforward comparison of trained model outputs, a consistent prediction time window was maintained in all the different setups, i.e., the last month of the year collected in the dataset (as shown in Figure 2).

1) MARKET INFLUENCE

To analyze different markets' regions, distinguishing and measuring the effects of geographical and economic factors on the stock predictions, the banks affiliated with the Federal Reserve and the US federal government, and the ones affiliated with the European Central Bank were separated.

2) BANKS AGGREGATION

By aggregating several banks together, the aim was to assess whether employing data from more banks could help the model learn more general patterns independently from the given bank. Thus, *local* and *global* experiments were performed. In the former case, one model was trained for each bank. In the latter, the training and test sets were built by aggregating the data from all the banks available. However, to avoid potential noises given by banks belonging to different markets, the territory division was maintained, including either only the banks related to the USA or the EU economies. The evaluation was made on a test set encapsulating all the test samples of the individual banks from that economic area.

3) TRAINING TIME SPAN

The time span γ related to training data can have an impact on performance. In general, the availability of the training data is a crucial factor in training predictive models effectively, especially for deep learning approaches. However, including excessively old data in the training set can introduce some challenges, since financial market data is highly dynamic and subject to various factors such as economic events, policy changes, and market trends, specific to certain time periods. Older data may thus contain patterns and trends that are no longer relevant or reflective of the current market dynamics. To take this aspect into account, the benefits of training the

[5] https://www.trackmyhashtag.com/

[6] A detailed information on these profiles may be found in Appendix B.

[7] https://apify.com/

[8] https://pypi.org/project/langdetect/

[9] https://www.nytimes.com/games/wordle/index.html

[10] Since FinBERT is trained for a sentiment analysis task in a domain akin to the one presented by the collected tweets, the model was implemented without any adaptation and fine-tuning steps to the work use case. This avoids having an expensive annotation phase for the tweets.

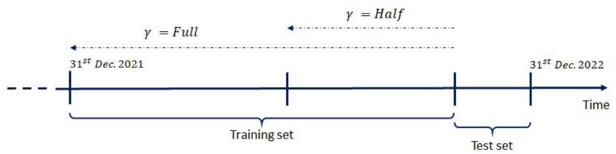

FIGURE 2. Training/test split with different values of γ.

models with a time span γ either *full* or *half* of dataset, as shown in Figure 2, were explored.

In the former case, data from *December,* 31st *2021* to *November,* 31st *2022* were employed for the training phase of the model, while in the second scenario data from *July,* 31st *2022* were used, again, up to the last month (December 2022), which was employed for testing for all the experimental setups.

4) SEQUENCE TIME SPAN

Long sequences potentially provide the model with more contextual information and allow it to capture longer-term dependencies in the data. However, there are trade-offs associated with using longer sequences, especially in domains like the stock market where the data can be volatile and noisy, making it more challenging to extract meaningful patterns and trends. On the other hand, shorter sequences may provide more focused and immediate information to the model, helping in the capture of more recent trends and patterns. To explore the impact of sequence length β on model performance, three different lengths were experimented with: *5 days* (one stock market week), *10 days* (two weeks), and *20 days* (one stock market month).

5) SENTIMENT CONTRIBUTION

What can be said about the contribution of the extracted sentiment features? To assess this aspect, different experiments with different features sets were conducted, either with (a total of $n = 8$ features) or without (only the $n = 5$ stock features) the sentiment triplets gathered from Twitter with FinBERT. By considering the two distinct markets, here the aim was also to explore how sentiment analysis on Twitter can vary across different geographical regions and markets.

D. EVALUATION

At the outset of the proposed analysis, the Pearson correlation index was applied to evaluate the relationships among various banks, distinctly categorizing them into US and EU markets. Then, to evaluate the presented approach and take into account the limitations given by the imbalanced datasets, the F_1-score was calculated. Such a metric harmonizes precision and recall, offering a balanced measure of model performance across varying class distributions. The application of the F_1-score weighted metric is particularly critical, as it adjusts for the prevalence of classes within the test set, ensuring that each class's impact on the overall performance is proportional to its frequency. This methodological choice addresses the potential overestimation of model performance that can occur in unbalanced datasets.

IV. RESULTS AND DISCUSSION

In this study, several factors that could potentially have an impact on the development of a predictive model for stock prices were investigated. The results obtained by varying all these parameters are reported in Table 1.

A. MARKET INFLUENCE

As for the first result of the analysis, the obtained correlation results revealed a significantly high average value of 0.778 for US banks, with a peak correlation of 0.982 observed between Goldman Sachs and Morgan Stanley. This high correlation among US banks suggests a synchronized market behavior heavily influenced by unified economic policies from the Federal Reserve and the US federal government, which tend to promote a cohesive financial environment. Conversely, the correlation values for European banks stood at a lower average of 0.596. This reflects a greater diversity within the European banking market, attributable to the unique economic conditions prevailing in different EU member states. Despite the overarching regulations of the European Central Bank, the varied fiscal and monetary policies at the national level contribute to a less coherent inter-bank relationship.

This foundational diversity in the European market elucidates why European banks tend to perform better with aggregated data in a global approach. The lower intra-market correlation suggests that combining data from multiple banks can provide a more comprehensive understanding of the market dynamics, smoothing out idiosyncrasies of individual banks. It enables predictive models to capture broader economic trends rather than being overly swayed by the anomalies of single entities. In essence, the aggregation helps to mitigate the noise introduced by the diverse economic policies across Europe, leading to more stable and reliable predictive performance.

The higher inter-bank correlation observed in the US market, compared to the more fragmented European market, suggests that the management strategies should be tailored to the specific market conditions. In the US, unified economic policies lead to synchronized market behaviors, making it feasible to apply more generalized management approaches. In contrast, the diversity within the European market necessitates a more nuanced approach that considers the varied economic conditions across different countries.

The classification results, as detailed in Table 1, illustrate significant differences between the North-American and European banking sectors, too. For North-American banks, the models achieved an encouraging average F_1-score of 0.68, with some banks reaching scores around 0.80. This indicates a strong performance despite the inherent volatility of the banking industry during the studied period. In contrast, European banks exhibited lower effectiveness,

TABLE 1. Performance (in terms of F_1-scores) of the best settings found across American (top block) and European (bottom block) banks within the classification task. The banks are reported (by their acronyms) in order of their market capitalization.

Area	Bank	F_1-score	Banks aggregation	Training span (γ)	Sentiment contribution	Sequence span (β)
USA	JPM	0.542	Local	Half	Yes	20
	BAC	0.800	Global	Half	Yes	5
	TD	0.500	Global	Full	Yes	10
	MS	0.664	Local	Full	Yes	10
	GS	0.707	Global	Half	No	5
	BMO	0.806	Global	Half	No	5
	PNC	0.795	Global	Half	No	5
	USB	0.600	Global	Half	Yes	10
	TFC	0.700	Global	Full	No	5
	COF	0.797	Global	Half	Yes	5
	MTB	0.600	Global	Half	No	10
	FITB	0.791	Global	Full	No	5
	RF	0.560	Local	Half	No	10
	NTRS	0.645	Global	Full	Yes	10
	KEY	0.680	Global	Half	Yes	5
	HBAN	0.762	Local	Full	No	10
	ALLY	0.649	Local	Half	No	10
EU	BNPQF	0.181	Global	Full	No	10
	UBS	0.450	Global	Half	No	5
	SAN	0.543	Global	Full	No	5
	ING	0.539	Global	Half	No	5
	IITFS	0.512	Global	Full	No	10
	BBVA	0.450	Global	Full	No	5
	UNCFF	0.576	Global	Full	No	20
	CRARF	0.279	Global	Full	No	10
	DB	0.555	Global	Half	No	5

with an average F_1-score of 0.447 and a minimum score of 0.181 for BNP Paribas, signaling the complexities of accurately predicting stock movements in these markets.

B. BANKS AGGREGATION

Interestingly, for most of the banks, the global approach, in which the banks were aggregated within the given market, was particularly successful when compared to its local counterpart, in which the models were trained using only the data pertaining to the individual bank under investigation, especially in Europe. This observation underscores the value of a comprehensive dataset in reflecting broader market trends and improving model robustness. Especially for the European market, for which the global setting was predominant. Noteworthy, among the US area, the few banks showing better performances with a local setting, i.e., employing only the data coming from the given bank, are in the top and bottom groups per market capitalization. This may indicate that "out-of-scale" assets may diverge from the collective behavior of the market.

This aggregation can thus help to mitigate the noise introduced by diverse economic policies and provides a more comprehensive understanding of market dynamics, thus improving the reliability of risk management models.

C. TRAINING AND SEQUENCE TIME SPANS

By analyzing the training time span γ, a preference for models trained solely on the last half of the dataset was observe, which aligns with the expectations that recent trends are more indicative of future movements. However, this trend does not uniformly apply across all banks and regions, indicating the importance of market-specific strategies and the potential value of incorporating longer historical data periods in certain cases. Indeed, for both areas, the best performances were obtained by exploiting a short sequence time span (β), i.e., 1 or 2 market stock weeks, which have been outperformed by the 1-month setting in only few scenarios. Particularly only one for the US area (for JP Morgan) and one for Europe (for UniCredit). Nonetheless, the diversity in sequence span selection might reflect different banking strategies and the significance of capturing various time-dependent patterns.

D. SENTIMENT CONTRIBUTION

Regarding the role of sentiment analysis, the findings indicate that incorporating Twitter sentiment data into the models offers mixed benefits. While it appears to enhance classification performance for some of the largest North-American banks by market capitalization, it does not uniformly improve outcomes across all banks, particularly in Europe.

This variability may be attributed to the geographic focus of the sentiment data and the methods used for sentiment aggregation.

In the presented cases, the integration of sentiment analysis from social media platforms can provide additional insights into market movements driven by public sentiment. This approach can be particularly useful for risk managers to anticipate market reactions to news and events in real-time. However, the employed Twitter dataset comprises tweets from selected influential financial analysts and CEOs, and while this provides valuable sentiment data, it may not fully represent the broader market sentiment, leading to potential biases.

V. CONCLUSION

The study on predictive modeling for bank stock prices in the US and Europe has revealed that market-specific dynamics significantly influence model performance. The higher inter-bank correlation in the US suggests that unified economic policies lead to more synchronized market behaviors, supporting models based on individual bank data. In contrast, the greater diversity among European banks, due to varied national economic policies, indicates that aggregated data models are more effective, capturing broader market trends and smoothing out individual discrepancies. Additionally, models trained on recent data generally outperform those based on longer historical periods, though the optimal data span varies by region and bank. Incorporating Twitter sentiment data has proven beneficial primarily for North American banks.

Limitations of this study include the exclusive focus on bank stocks. Future research should expand the scope to include other stock groups, particularly those that are less interconnected, to evaluate different market contexts. Additionally, comparing these findings with other classification (and sentiment analysis) models or techniques, such as a regression-to-classification approach, would provide a valuable additional dimension of analysis. Nevertheless, this study highlights the importance of tailoring predictive models to reflect regional economic conditions and the potential of integrating diverse data sources to enhance forecasting accuracy, providing a valuable resource for researchers, especially newcomers to this intricate field.

APPENDIX A MODEL ARCHITECTURE

As illustrated in Figure 3, the employed models are built using several layers arranged in a specific sequence:

- An input layer that receives a $\beta \times n$ matrix for each sample. The matrix size is determined by the number of β timestamps (referred to in Section III-C4) and the number of features n for each timestamp (see Section III-C5)
- A series of n_{lstm} LSTM layers, consisting of n_i hidden units (with $i \in [1 - n_{lstm}]$) with a_{lstm} activation function

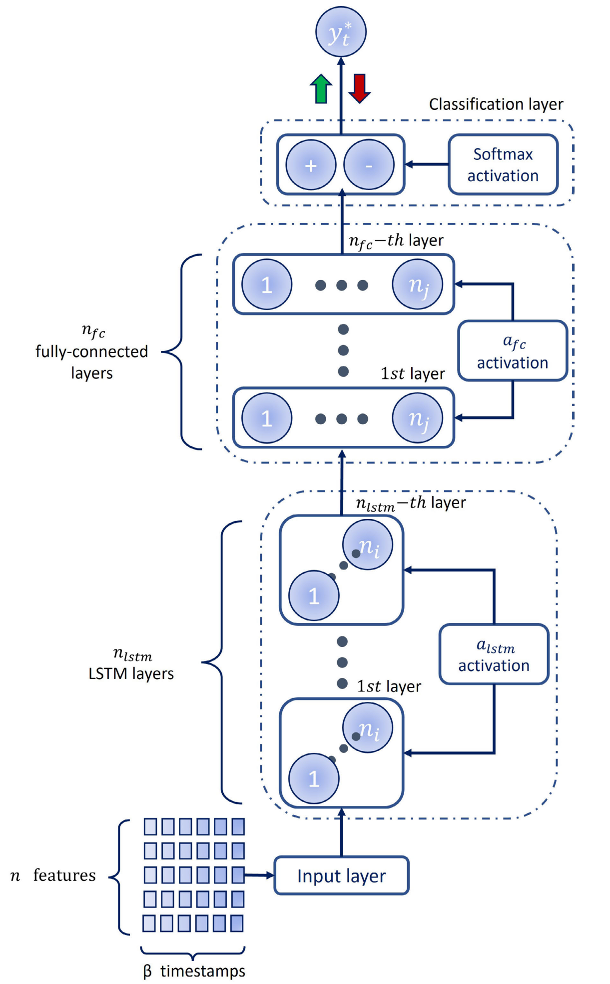

FIGURE 3. Model architecture.

- Several n_{fc} stacked fully-connected layers, consisting of n_j hidden units (with $j \in [1 - n_{fc}]$) with a_{fc} activation function
- One final output layer, consisting of two neurons (fully-connected to the representation extracted from the previous last fully-connected layer) and a softmax activation, which provides prediction probabilities for the positive and negative trend classes

Different combinations of these layers and settings were tested. For the LSTM and fully-connected layers, various numbers of layers and units were tried, using either ReLU or tanh activation functions. The learning rate and batch size for training were also varied. The best setup involved two LSTM layers with specific numbers of hidden units and ReLU activation, followed by two dense layers. This model used the Adam optimizer.

APPENDIX B TWITTER ACCOUNTS

For the sake of completeness, the Twitter accounts included in this study are listed in Table 2, together with information

TABLE 2. The Twitter accounts selected for this study, reported by their username, together with their description, the number of followers and the Tweets-per-Month (TpM) ratio averaged across the twelve months included in the acquired dataset.

Username	TpM	Followers (K)	Description
@tradertvneal	66.42	27.7	Trader and host of traderTVLIVE Living the dream, trading the markets
@scottwapnerCNBC	66.25	251.6	Host, CNBC's Halftime Report and Closing Bell
@paulkrugman	65.17	4500.0	Nobel laureate. Op-Ed columnist, @nytopinion
@Burns277	62.67	79.5	Trader - Angel Investor- Alt Investments - Crypto Investor
@jimcramer	60.00	2000	Host of madmoneyoncnbc and I run the CNBC Investing Club
@CarbonTracker	58.75	50.2	Financial specialists making climate risk visible to financial markets.
@ronshevlin	58.42	41.2	Chief Research Officer at Cornerstone Advisors. Senior Contributor at Forbes
@jeffcoxcnbccom	56.00	9.5	CNBC econ editor, wrote a couple books
@elerianm	53.33	630.3	President, Queens' College
@msci	53.25	25.4	An independent provider of research-driven insights
@ritholiz	52.25	213.0	Chair/CIO of RWM http://ritholtzwealth.com
@michaelsantoli	50.92	102.0	CNBC Senior Markets Commentator.
@awealthofcs	49.08	233.8	Trying to bring some common sense to the world of finance
@byHeatherLong	47.17	68.4	WashingtonPost economic columnist and Editorial Board member
@FTSERussel	44.67	18.5	FTSE Russell is a leading global provider of benchmarks, analytics, and data solutions with multi-asset capabilities.
@tracyalloway	43.58	185.8	Co-host of the Odd Lots podcast
@crescatkevin	40.92	35.3	@Hedge fund manager: global macro, long/short equity, precious metals
@rbccm	39.50	4.5	The official Twitter account of RBC Capital Markets.
@CatoInstitute	37.17	366.8	A public policy research organization dedicated to individual liberty, limited government, free markets, and peace.
@Sruhle	36.33	809.6	Host of 11thhour on msnbc M-F 11PM EST, nbcnews Snr Business Analyst
@goldmansachs	35.25	995.4	We are Goldman Sachs
@joesquawk	30.83	218.6	Official Twitter account for Squawk Box's Joe Kernen.
@rinsana	26.92	24.3	I am a contributor to CNBC, public speaker, and the Co-CEO of Contrast Capital Partners.
@marklehman	18.92	2.5	Morning reporter for News 6
@lizmoyer	18.00	3.4	Barron's editor. Alum of CNBC, NYTimes Dealbook. WSJ, Dow Jones, and Forbes.
@timcook	15.25	14000.0	CEO APPLE
@sundarpichai	13.50	52000.0	CEO, Google and Alphabet
@michaelmiebach	12.75	7.3	CEO MASTERCARD
@thesalwart	9.75	342.6	Co-host of the Odd Lots podcast
@satya-nadella	9.00	3000.0	Chairman and CEO of Microsoft Corporation
@peterbrandt	7.50	696,9	Prop trader since 1975. Author. Featured in Market Wizard Series
@ginachon	6.08	5.5	Biz & Tech Editor Semafor in DC
@AndyJassy	5.75	157.5	Lead Amazon
@Buffet	2.92	1600.0	Chairman and CEO of Berkshire Hathaway

regarding the type of account, the number of followers, and the ratio of tweets per month during the span period under analysis.

AUTHORS' CONTRIBUTION

Luca Bacco: Methodology, Software, Investigation, Formal Analysis, Visualization, Writing - Original Draft. **Lorenzo Petrosino:** Methodology, Software, Formal Analysis, Investigation, Writing - Original Draft. **Domenico Arganese:** Software, Data Curation. **Marco Papi:** Conceptualization, Writing - Original Draft, Validation, Investigation. **Luca Vollero:** Conceptualization, Writing - Review & Editing, Validation, Supervision. **Mario Merone:** Conceptualization, Methodology, Validation, Visualization, Writing - Original Draft, Supervision, Project Administration.

REFERENCES

[1] B. Weng, M. A. Ahmed, and F. M. Megahed, "Stock market one-day ahead movement prediction using disparate data sources," *Expert Syst. Appl.*, vol. 79, pp. 153–163, Aug. 2017.
[2] M. S. Checkley, D. A. Higón, and H. Alles, "The hasty wisdom of the mob: How market sentiment predicts stock market behavior," *Expert Syst. Appl.*, vol. 77, pp. 256–263, Jul. 2017.
[3] T. M. Nisar and M. Yeung, "Twitter as a tool for forecasting stock market movements: A short-window event study," *J. Finance Data Sci.*, vol. 4, no. 2, pp. 101–119, Jun. 2018.
[4] J. A. Ou and S. H. Penman, "Financial statement analysis and the prediction of stock returns," *J. Accounting Econ.*, vol. 11, no. 4, pp. 295–329, Nov. 1989.
[5] E. F. Fama, "The behavior of stock-market prices," *J. Bus.*, vol. 38, no. 1, p. 34, Jan. 1965.
[6] E. F. Fama, "Random walks in stock market prices," *Financial Analysts J.*, vol. 51, no. 1, pp. 75–80, Jan. 1995.
[7] J. C. Van Horne and G. G. C. Parker, "The random-walk theory: An empirical test," *Financial Analysts J.*, vol. 23, no. 6, pp. 87–92, Nov. 1967.

[8] M. C. Jensen and G. A. Benington, "Random walks and technical theories: Some additional evidence," *J. Finance*, vol. 25, no. 2, p. 469, May 1970.

[9] O. Bustos and A. Pomares-Quimbaya, "Stock market movement forecast: A systematic review," *Expert Syst. Appl.*, vol. 156, Oct. 2020, Art. no. 113464.

[10] S. Erfanian, Y. Zhou, A. Razzaq, A. Abbas, A. A. Safeer, and T. Li, "Predicting Bitcoin (BTC) price in the context of economic theories: A machine learning approach," *Entropy*, vol. 24, no. 10, p. 1487, Oct. 2022.

[11] X. Zhong and D. Enke, "A comprehensive cluster and classification mining procedure for daily stock market return forecasting," *Neurocomputing*, vol. 267, pp. 152–168, Dec. 2017.

[12] L. Guang, W. Xiaojie, and L. Ruifan, "Multi-scale RCNN model for financial time-series classification," 2019, *arXiv:1911.09359*.

[13] A. Kazem, E. Sharifi, F. K. Hussain, M. Saberi, and O. K. Hussain, "Support vector regression with chaos-based firefly algorithm for stock market price forecasting," *Appl. Soft Comput.*, vol. 13, no. 2, pp. 947–958, Feb. 2013.

[14] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 2327–2333.

[15] Q. Wen, Z. Yang, Y. Song, and P. Jia, "Automatic stock decision support system based on box theory and SVM algorithm," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1015–1022, Mar. 2010.

[16] S. Chen and H. He, "Stock prediction using convolutional neural network," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 435, Nov. 2018, Art. no. 012026.

[17] Y. Liu, Q. Zeng, H. Yang, and A. Carrio, "Stock price movement prediction from financial news with deep learning and knowledge graph embedding," in *Proc. 15th Pacific Rim Knowl. Acquisition Workshop (PKAW)*, Nanjing, China. Cham, Switzerland: Springer, Aug. 2018, pp. 102–113.

[18] A. K. Saxena, R. R. Dixit, and A. Aman-Ullah, "An LSTM neural network approach to resource allocation in hospital management systems," *Int. J. Appl. Health Care Anal.*, vol. 7, no. 2, pp. 1–12, 2022.

[19] Y. Ju, "LSTM-based volatility prediction study of SSE index," in *Proc. 5th Int. Conf. Mechatronics Comput. Technol. Eng. (MCTE)*, Dec. 2022, pp. 1173–1178.

[20] A. Vidal and W. Kristjanpoller, "Gold volatility prediction using a CNN-LSTM approach," *Expert Syst. Appl.*, vol. 157, Nov. 2020, Art. no. 113481.

[21] E. Koo and G. Kim, "A hybrid prediction model integrating GARCH models with a distribution manipulation strategy based on LSTM networks for stock market volatility," *IEEE Access*, vol. 10, pp. 34743–34754, 2022.

[22] W. Wang and Y. Wu, "Risk analysis of the Chinese financial market with the application of a novel hybrid volatility prediction model," *Mathematics*, vol. 11, no. 18, p. 3937, Sep. 2023.

[23] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 9713–9729, Jul. 2020.

[24] S. Gite, H. Khatavkar, K. Kotecha, S. Srivastava, P. Maheshwari, and N. Pandey, "Explainable stock prices prediction from financial news articles using sentiment analysis," *PeerJ Comput. Sci.*, vol. 7, p. e340, Jan. 2021.

[25] K. Du, F. Xing, R. Mao, and E. Cambria, "Financial sentiment analysis: Techniques and applications," *ACM Comput. Surv.*, vol. 56, no. 9, pp. 1–42, Oct. 2024.

[26] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in *Proc. IEEE 5th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Apr. 2019, pp. 205–208.

[27] G. Bathla and S. Gupta, "FinBERT and LSTM-based novel model for stock price prediction using technical indicators and financial news," *Int. J. Econ. Bus. Res.*, vol. 28, no. 1, pp. 1–16, 2024.

[28] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," in *Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst. (SCOPES)*, Oct. 2016, pp. 1345–1350.

[29] T. Swathi, N. Kasiviswanath, and A. A. Rao, "An optimal deep learning-based LSTM for stock price prediction using Twitter sentiment analysis," *Appl. Intell.*, vol. 52, no. 12, pp. 13675–13688, Sep. 2022.

[30] T. T. P. Souza, O. Kolchyna, P. C. Treleaven, and T. Aste, "Twitter sentiment analysis applied to finance: A case study in the retail industry," 2015, *arXiv:1507.00784*.

[31] M. S. Islam, M. N. Kabir, N. A. Ghani, K. Z. Zamli, N. S. A. Zulkifli, M. M. Rahman, and M. A. Moni, "Challenges and future in deep learning for sentiment analysis: A comprehensive review and a proposed novel hybrid approach," *Artif. Intell. Rev.*, vol. 57, no. 3, p. 62, Mar. 2024.

[32] T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *J. Finance*, vol. 66, no. 1, pp. 35–65, 2011.

[33] N. Oliveira, P. Cortez, and N. Areal, "Stock market sentiment lexicon acquisition using microblogging data and statistical measures," *Decis. Support Syst.*, vol. 85, pp. 62–73, May 2016.

[34] S. Sohangir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar, "Big data: Deep learning for financial sentiment analysis," *J. Big Data*, vol. 5, no. 1, pp. 1–25, Dec. 2018.

[35] D. Araci, "FinBERT: Financial sentiment analysis with pre-trained language models," 2019, *arXiv:1908.10063*.

[36] A. H. Huang, H. Wang, and Y. Yang, "FinBERT: A large language model for extracting information from financial text," *Contemp. Accounting Res.*, vol. 40, no. 2, pp. 806–841, May 2023.

[37] Y. Zhang and H. Zhang, "FinBERT–MRC: Financial named entity recognition using BERT under the machine reading comprehension paradigm," *Neural Process. Lett.*, vol. 55, no. 6, pp. 7393–7413, Dec. 2023.

[38] D. Egloff, M. Leippold, and P. Vanini, "A simple model of credit contagion," *J. Banking Finance*, vol. 31, no. 8, pp. 2475–2492, Aug. 2007.

[39] K. N. Naoui, "A dynamic conditional correlation analysis of financial contagion: The case of the subprime credit crisis," *Int. J. Econ. Finance*, vol. 2, no. 3, pp. 85–96, Jul. 2010.

[40] Z. Mighri and F. Mansouri, "Dynamic conditional correlation analysis of stock market contagion: Evidence from the 2007-2010 financial crises," *Int. J. Econ. Financial Issues*, vol. 3, no. 3, pp. 637–661, 2013.

[41] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

LUCA BACCO received the Ph.D. degree in information technology for biomedicine from the University Campus Bio-Medico di Roma, in 2023. He is currently a Biomedical Engineer. His expertise and current research interests include the artificial intelligence field, focusing on natural language processing and machine/deep learning for applications in healthcare.

LORENZO PETROSINO received the master's degree in biomedical engineering from the University Campus Bio-Medico di Roma, in 2020, where he is currently pursuing the Ph.D. degree in science and engineering for humans and the environment. His current research interests include machine and deep learning, edge computing, DLT and its interaction with the IoT, and decision support systems for diabetes.

DOMENICO ARGANESE is currently an Intelligent Systems Engineer with the University Campus Bio-Medico di Roma, in 2023. His expertise and current research fall into programming, intelligent systems simulation, and artificial intelligence applications in economics, finance, blockchain, and cybersecurity.

LUCA VOLLERO (Member, IEEE) is currently an Associate Professor of computer science with the University Campus Bio-Medico di Roma. His research interests include signal processing and digital imaging, AI, distributed systems, and embedded systems optimization. He is a member of the IEEE Computer Society, the IEEE Communication Society, ACM, and SIAM. He is a Co-Founder of the innovative startup Heremos, operating in the field of remote monitoring of frail persons and patients.

MARCO PAPI received the M.D. degree in mathematics from the University of Rome Tre, Italy, and the Ph.D. degree in mathematics from the University of Rome Tor Vergata. In 2008, he joined the Faculty of Engineering, University Campus Bio-Medico of Rome, as an Associate Professor of mathematical methods for financial applications. He is also an Adjunct Professor with the Department of Economics and Finance, University Luiss Guido Carli of Rome. From 2007 to 2014, he was a member of the Steering Committee and a Program Coordinator of the research network on advanced mathematical methods for finance, promoted by European Science Foundation (ESF).

MARIO MERONE (Member, IEEE) received the Ph.D. degree in bioengineering and bioscience, in 2017. He is currently an Assistant Professor with the University Campus Bio-Medico di Roma. He has extensive experience in the field of artificial intelligence (machine learning and deep learning for time-series analysis). His research has led to 40 publications, all focused on AI for healthcare. He is currently the scientific leader of several projects involving the application of artificial intelligence in the medical field. He is the Founding Partner of an innovative start-up called BPCOMEDIA S.R.L., focused on technology transfer of research results. He is a member of the IEEE Computer Society, the IEEE Computational Life Sciences, and the IEEE Sensors Council.

• • •

Open Access funding provided by ‘Università ''Campus Bio- Medico'' di Roma’ within the CRUI CARE Agreement