WORLD COMET RESEARCH GROUP

International Journal of Commerce, Finance and Digital Economy
Vol. 6, No. 2, 2023
Doi: 10.67228/3071642X/IJCFDE-2023PII9C2X
PP. 01-12

Original Article

Sentiment Analysis for Financial Market Prediction

Alan Bundy[1], Karen Spärck Jones[2]

[1]*Professor of Artificial Intelligence, University of Edinburgh, United Kingdom.*

[2]*Professor, University of Cambridge, United Kingdom.*

Received: 11-07-2023 Revised: 11-27-2023 Accepted: 12-01-2023 Published: 12-03-2023

ABSTRACT

Financial markets are highly sensitive to investor sentiment, news events, social media discussions, and macroeconomic announcements. Traditional market prediction models primarily rely on historical price and volume data, often overlooking the behavioral aspects that significantly influence market movements. This paper presents a comprehensive sentiment analysis framework for financial market prediction using natural language processing and machine learning techniques. Financial news articles, social media posts, and corporate announcements are collected and preprocessed using tokenization, stop-word removal, stemming, and feature extraction methods such as TF-IDF and word embeddings. Sentiment scores are generated using lexicon-based and deep learning approaches, and these scores are integrated with market indicators to predict stock price direction. Experimental evaluation demonstrates that combining textual sentiment with technical indicators improves prediction performance compared with price-based models alone. The proposed approach achieves higher accuracy, precision, and recall in forecasting short-term market trends. The study highlights the importance of investor psychology in financial forecasting and demonstrates how sentiment-driven models can assist traders, portfolio managers, and financial institutions in making informed investment decisions. The proposed framework provides a scalable and data-driven solution for intelligent financial market prediction.

KEYWORDS

Sentiment Analysis, Financial Market Prediction, Machine Learning, Natural Language Processing, Stock Market Forecasting, Deep Learning, Investor Sentiment.

ISSN: 3071-642X
PP: 01-12.

1. INTRODUCTION

1.1. Background

The working of financial markets is affected by many interrelated factors, such as the economic indicators, corporative earnings, government policy, geo-political events, market trends, investor psychology, etc. The task of predicting stock price movements is difficult, as the markets are continually changing and may be influenced by both quantitative financial information and qualitative information. The digital explosion has created a huge amount of financial data in recent years in the form of online news sites, social media sites, blogs, discussion boards and from official corporate communication. This information has a significant impact on investor sentiment and trading decisions and leads to immediate fluctuations in the prices of shares. This has made sentiment analysis a useful method for gaining insight into the emotional atmosphere and public opinion included in financial text. In the same way that researchers and financial analysts can draw conclusions from positive, negative and neutral sentiments from a variety of information sources, they can learn things about market behavior. By combining sentiment analysis with conventional financial data, investors, analysts, and financial institutions can make more informed decisions and better manage investment risks, helping to make more accurate stock market predictions.

1.2. Needs of Sentiment Analysis for Financial Market Prediction

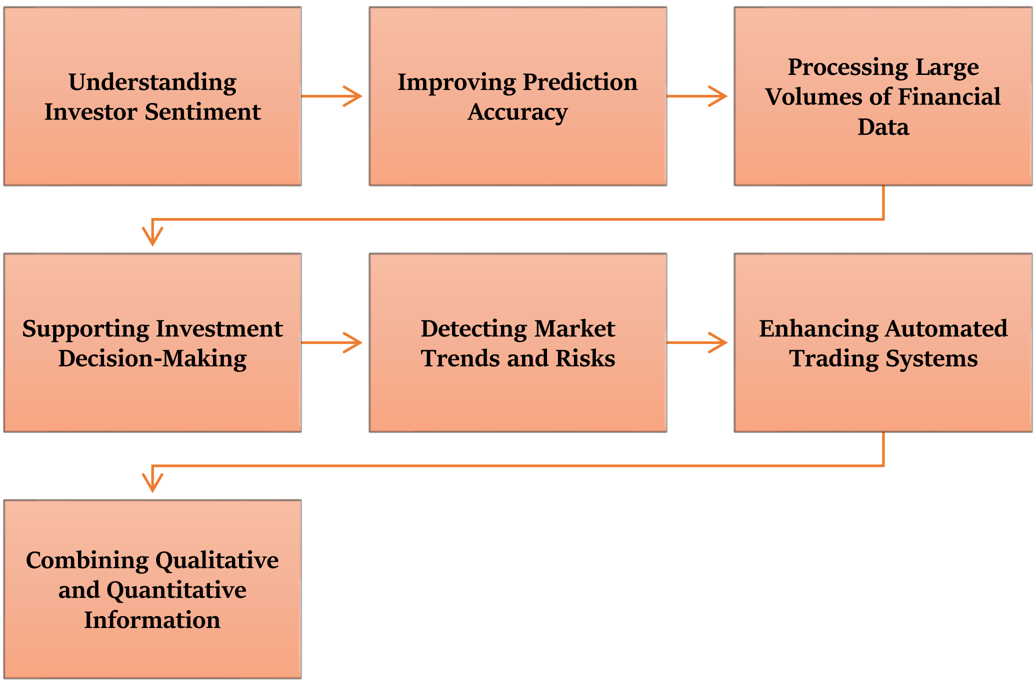

Fig 1 - Needs of Sentiment Analysis for Financial Market Prediction

1.2.1. Understanding Investor Sentiment

Investor sentiment is an important factor in driving movements in the stock market. When positive news is reported, positive opinions are expressed or good corporate news is given, investors tend to become more confident as a result, leading to increased buying activity, and stock prices rise. Likewise, bad news, financial reports, or negative economic events can cause a loss of confidence and fear and subsequently cause a market decline. These emotional reactions can be detected in financial text using sentiment analysis, which can improve market predictions.

1.2.2. Improving Prediction Accuracy

Traditional methods for predicting the stock market are mostly based on past stock price data and indicators. But, such models can miss sudden changes in the market, such as breaking news or shifting investor sentiment. Sentiment Analysis provides qualitative data that can be used to refine prediction models, resulting in more accurate and reliable forecasts for market trends.

1.2.3. Processing Large Volumes of Financial Data

Digital platforms are generating millions of financial news articles, social media posts, blogs, and company statements every day. When it comes to manually processing this kind of information, it's time consuming and impractical. Sentiment analysis is a technique that can be used to extract valuable information from these unstructured data sources, enabling investors and analysts to handle information efficiently and respond promptly to market movements.

1.2.4. Supporting Investment Decision-Making

Sentiment analysis gives investors, portfolio managers and financial institutions the information they need to make informed investment decisions. When investors understand the positive and negative market sentiment early, they will be able to make the right decision on their buying, selling, or holding strategies and ensure that the investment will not pose any risk but, rather, maximize the returns.

1.2.5. Detecting Market Trends and Risks

Economic events, government policies, and global occurrences have a significant impact on financial markets. By gauging shifts in public opinion and investor confidence, sentiment analysis helps identify market trends before they materialize, allowing for proactive adjustments to strategies. Sentiment analysis can help detect shifts in the market before they happen, as it can help gauge public opinion and investor confidence. This enables the financial analysts to identify market opportunities as well as any risks that are coming through before the stock price.

1.2.6. Enhancing Automated Trading Systems

The modern algorithmic and automated trading systems need to operate on the information available in real-time to make trading decisions efficiently. Financial news, corporate disclosures, and social media discussions can be continuously monitored to provide real-time sentiment signals using sentiment analysis. These indicators can be combined with technical indicators to enhance the effectiveness of trading strategies, leading to quicker and more precise reactions to market developments.

1.2.7. Combining Qualitative and Quantitative Information

Numerical financial indicators, along with qualitative data like investors' confidence, company reputation, and opinions have an influence on the movements of the stock market.

Sentiment analysis can help fill this void by quantifying the sentiment into a sentiment score from the text, which can then be added to the technical indicators. This holistic perspective helps to better understand market behavior and enhance the performance of financial market prediction models.

1.3. Problem Statement

Due to the volume, dynamism and uncertainty of financial markets, predicting stock market movements is one of the most difficult problems in financial analysis. Historical stock prices, trading volume, and technical indicators are the key components of traditional prediction methods, which are used to predict market trends. While these approaches work well at charting historical trends, they tend not to factor in other events that can lead to sudden upsets in stock prices, like breaking financial news, corporate announcements, government policies, geopolitical events, and investor emotions. Concurrently, the digital media has given birth to an overwhelming amount of financial information every day, on news sites, social media platforms, blogs, discussion forums, and official company reports. The data sources are very diverse, and encompass structured and unstructured data in various forms, with writing styles and levels of trustworthiness. Obtaining meaningful sentiment from these varied types of text is difficult as financial terminology is context-sensitive, with domain-specific terms, abbreviations, and expressions that can be taken out of context. In addition, the text content on social media can often contain informal language, slang, abbreviations, emojis, and potentially misleading or noisy information, complicating sentiment classification even more. Another difficulty is to effectively combine the textual sentiment with conventional market metrics like stock prices, trading volume, moving averages, Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD). These qualitative and quantitative characteristics need to be merged and advanced machine learning and deep learning models need to be developed that can learn the intricate connections between investor sentiment and market performance. Furthermore, there is a high volatility in financial markets and the impact of sentiment on stock prices will differ between companies, industry and market. Thus, it is necessary to develop a strong prediction model to accurately quantify financial sentiment, remove noise signals in text information and use sentiment information to combine technical indicators to comprehensively enhance the prediction accuracy. Solving these challenges would result in more precise stock market predictions, aid in financial decision making, financial risk reduction, and improve the effectiveness of intelligent systems in financial decision support.

2. *LITERATURE SURVEY*

2.1. Early Financial Sentiment Models

The initial financial sentiment analysis methods were based on dictionary approaches, which included lists of predefined positive and negative words used to classify the sentiment of financial documents. The most popular one is the Loughran–McDonald Financial Dictionary, which was designed for financial documents since it is challenging to correctly classify financial words in general-purpose sentiment dictionaries. For instance, negative words like "liability" or "depreciation" might seem negative in normal usage but are not in financial reporting. The advantage of the dictionary-based methods is that they are simple to implement, require limited training data and are computationally efficient. They have difficulty capturing context, sarcasm, negation and the different meanings of words in different financial situations, however, and so they are not very accurate at predicting the meaning of words.

2.2. Machine Learning Approaches

Lexicon-based methods have limitations, so machine learning algorithms have been developed for financial sentiment classification. Popular methods include Support Vector Machines (SVM), Naïve Bayes, Decision Trees and Random Forests that are trained on labelled datasets to learn the patterns of sentiment instead of using fixed dictionaries. These models rely on various linguistic features like term frequency, TF-IDF, and n-grams to detect positive, negative, or neutral sentiment within financial news and reports. Generally, machine learning methods are more accurate and more adaptable than dictionary-based methods, as they can identify complex relationships between words. However, they heavily rely on the availability of high-quality labeled datasets and feature engineering.

2.3. Deep Learning Methods

With the development of artificial intelligence, the deep learning model is adopted for financial sentiment analysis. The deep learning model is adopted in financial sentiment analysis due to recent developments in artificial intelligence. Contextual models like Long Short-Term Memory (LSTM) networks and transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) can learn from the relationships between words and long-distance context in financial documents. These methods learn meaningful representations from raw text automatically, unlike traditional machine learning models, where an extensive amount of manual feature extraction is required. Pre-trained transformer models can further be fine-tuned to provide a much better accuracy in sentiment classification for specific financial applications. As a result, numerous state-of-the-art financial forecasting models are now using deep learning techniques.

2.4. Research Gap

While financial sentiment analysis is well advanced, there are still challenges. Most of the current research concentrates on textual sentiment analysis from financial news, social media or company reports or only on the technical analysis of stock prices, trading volume and market trend. It is a separation that can reduce the predictive power of stock market forecasting models by affecting the market's behaviour due to investor sentiment and quantitative financial indicators. Hence, the need for a common platform that can effectively merge sentiment analysis with technical market data has been growing. This can have the effect of more accurate, robust and reliable stock price predictions which could be an important direction to future research.

3. METHODOLOGY

3.1. Data Collection

The most crucial part of any financial sentiment analysis and stock prediction system is data collection, and the success or failure of the prediction model depends on the quality and reliability of the data gathered. The data in this study comes from several sources of reliable data, so that the market is covered. Financial news articles are gathered from trusted business news sites, which offer up to-the-minute information around company performance, economic policies, mergers and acquisitions, earnings reporting, along with world financial occurrences that affect investor feeling. Finally, stock exchange announcements such as quarterly financial statements, dividend declaration, annual reports, regulatory disclosures etc, are also collected as they contain information that has been confirmed and released by listed companies. The company also monitors social media channels, including X (formerly Twitter), Reddit, and financial forums, to gauge public sentiment and investor reactions to financial news as these networks often provide real-time insights into market response and investor emotions. In addition to the text data, reliable financial databases or stock market APIs are used to retrieve historical stock market information like opening price, closing price, highest price, lowest price, adjusted closing price and daily trading volume. These numerical indicators are vital details regarding market trends and price fluctuations over time. The dataset is collected over a sufficient time frame that aligns with various market conditions such as bull markets, bear markets, and market stability, allowing the model to learn from different market behaviors. All datasets are synced using a time stamp after collection, so that news articles, social media posts and stock prices match the same trading period. A duplicate record, missing values and data that would be irrelevant to further processing is removed to enhance data quality. The dataset combines text data with historical market data, creating a comprehensive representation of factors that can affect stock prices. The multi-source data collection strategy provides a foundation for the proposed system to gather and analyze investor sentiment and quantitative market indicators, further enhancing the robustness, accuracy, and reliability of financial sentiment analysis and stock market prediction.

3.2. Text Preprocessing

In financial sentiment analysis, preprocessing the text is an important phase, as the raw data gathered from news articles, company announcements, and social media platforms may be noisy, inconsistent, or include irrelevant information that can impact the effectiveness of machine learning and deep learning models. The first step of pre-processing aims to clean and normalize the text, to enable the text to be analyzed by sentiment classification algorithms effectively. The first step in the process is tokenization, in which every document or sentence is broken down into smaller segments known as tokens, typically words or meaningful phrases. The model can ingest text in a structured format using tokenization. Lowercasing is also performed after tokenization, to convert the entire word to lowercase, which ensures that words like "Profit", "PROFIT" and "profit" will be treated as the same word and will not increase the vocabulary needless. The next step in stop-word removal is the identification of words that appear too frequently, like Eliminating the words the, is, and, of, and to reduces the amount of data needed for sentiment classification, while also reducing the complexity of the calculations. After the stop words are deleted, stemming will be done to remove prefix or suffix of words to make them in their root form. For instance, the word "investing" could be shortened to "invest", "invested" could be shortened to "invested", and the word "investment" could be shortened to "investment". While stemming is fast, it sometimes yields partial or clumsily formed stem words. To address this problem, an additional processing step, namely lemmatization, is also employed to transform words into their meaningful base forms, based on vocabulary and grammatical analysis. For example, ‘better' is changed to ‘good' and ‘running' to ‘run' while

ISSN: 3071-642X
PP: 01-12.

maintaining the meaning of the text. Other pre-processing operations could be the removal of punctuation marks, special symbols, URLs, hashtags, emojis, numbers, and/or removing repeated entries, and correcting spelling mistakes if needed. These preprocessings techniques ensure the consistency of the data, reduce its dimensionality, and remove noise from the data. This in turn ensures that the processed text is more informative, more appropriate for feature extraction and sentiment classification, and thus improves the accuracy, efficiency and the overall performance of the proposed financial sentiment analysis and stock market prediction model.

3.3. Feature Extraction

In financial sentiment analysis, the process of feature extraction is crucial because computer science algorithms like machine learning or deep learning can not read raw text directly. Rather, the textual information should be mapped to numerical representations that maintain the meaning of the content in terms of both semantic and context. For the effective representation of financial text, two popular feature extraction methods are used in this study: the Term Frequency-Inverse Document Frequency (TF-IDF) method and word embeddings. TF-IDF is a statistical technique used to determine how significant a word is in a particular document, compared to a set of documents. It gives greater weight to words that occur often in one document but not so often in others, thus accentuating words with a high level of information content. For instance, the term “profit,” or “earnings” or “growth”, or “loss” or “bankruptcy” can be assigned greater weight if it is a unique characteristic of the sentiment of a financial news article. TF-IDF generates a sparse numerical feature matrix suitable for traditional machine learning algorithms like Support Vector Machines (SVM), Naive Bayes and Random Forests. TF-IDF however, fails to identify semantic relationship or context meaning between words. To overcome the above shortcoming, word embeddings are used. Word embedding methods like Word2Vec, GloVe and transformer-based embeddings produce dense representations in which semantically similar words are closer in a multidimensional vector space. This helps the model to grasp contextual similarities and relationships between financial terms, even if they have the same meaning in different words. For example, terms like "gain," "increase," and "growth" are assigned similar vector representations, due to the relationship between these different financial terms. Word embeddings can also help deep learning models like Long Short-Term Memory (LSTM) networks and Bidirectional Encoder Representations from Transformers (BERT) to keep the contextual information and enhance the sentiment classification accuracy. The proposed system leverages both statistical significance and semantic meaning of financial text, using TF-IDF and word embeddings. The hybrid feature extraction method yields more comprehensive and informative representations of textual information, resulting in better sentiment analysis, improved model learning, and more accurate stock market prediction.

3.4. Prediction Model

The prediction model is the main element of the proposed financial sentiment analysis system, which is used to integrate textual sentiment information and historical financial market data and make predictions about the future behavior of the stock market. Sentiment features are derived from financial news articles, company announcements, and social media content using methods like TF-IDF and word embeddings, followed by the incorporation of technical indicators based on past stock market information. Technical indicators are quantitative data related to market behavior, including features like opening price, closing price, highest price, lowest price, trading volume, moving averages, Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Bollinger Bands and others that follow the market's trend. By combining these two types of information, the model can better reflect the qualitative and quantitative factors that affect the changes in stock prices. This is then fed into a Long Short-Term Memory (LSTM) neural network. A specialized variant of recurrent neural networks (RNNs) that are better suited to process sequential and time-series data, and that avoid some of the issues that traditional RNNs face, such as the vanishing gradient problem. It has memory cells and gating mechanisms such as input, forget and output gates so that it can store significant historical data and forget the irrelevant data in long sequences. This is something that makes LSTM very well suited for stock market predictions, as the sentiment and behavior of the market in the past has a large impact on future movements in the market. In the training phase, the model identifies intricate patterns and connections among sentiment indicators, technical indicators, and stock price variations, with the objective of minimizing prediction errors. During training, the model captures complex relationships and patterns between sentiment signals, technical indicators, and stock price fluctuations by optimizing the parameters using methods like backpropagation and Adam. After training, LSTM network forecasts the direction of the market (bearish, bullish or neutral) using the new financial data and the latest market data. Evaluation is performed using standard evaluation metrics like when the task is classification: accuracy, precision, recall, F1-score, or when task is regression: Root Mean Square Error (RMSE). The proposed prediction model integrates sentiment analysis with technical indicators in LSTM, thereby offering a thorough insight into the market dynamics and boosting the accuracy of predictions, helping in better decision making, and more reliable stock market predictions.

4. RESULTS AND DISCUSSION

4.1. Experimental Setup

The proposed financial sentiment analysis and stock market prediction model were tested with the help of historical financial data, sentiment information, and technical indicators in the experimental setup. The data was comprised of financial news articles, company announcements, social media posts, and historical stock market information such as opening price, closing price, highest price, lowest price, adjusted closing price and trading volume. The data collected was preprocessed before training by cleaning and tokenizing the text, lowercasing it, removing stop words, stemming, lemmatizing, and extracting features using TF-IDF and word embeddings.

Technical indicators like moving averages, Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD) were also computed and were used together with the retrieved sentiment indicators to form a composite indicator. The dataset was then split into training and testing data sets, and the model was trained with around 80% of the data, with the rest 20% used to test and evaluate the model's performance. The LSTM network was used during the training period to learn the relationship between historical market actions, technical indicators, and investor sentiment, minimizing the prediction error using backpropagation and Adam optimization

ISSN: 3071-642X
PP: 01-12.

algorithm. The learning rate, batch size, number of hidden layers, number of LSTM units, dropout rate, and the number of epochs for training were optimized to achieve a high level of performance and minimize overfitting. The unseen test set was used for the post-training assessment of the model with a set of standard performance metrics.

The accuracy was defined as the percentage correctly predicted market directions; precision was defined as the percentage of correctly identified positive predictions of market directions among all positive predictions of market directions. The recall rate was used to evaluate the performance of the model in detecting all the relevant positive cases, while the F1 score, which is the harmonic mean of precision and recall, was used to get a balanced evaluation of the classification performance, especially in the case of imbalanced data. These assessment measurements provided a complete picture of the model's forecasting precision and trustworthiness. The design of the experimental setup ensured that the proposed approach was fairly and systematically evaluated to prove its efficacy in combining financial sentiment analysis and technical indicators for precise and powerful stock market forecasting.

4.2. Performance Table

Table 1: Performance Table

Metric	Value
Accuracy	87.5
Precision	85.2
Recall	86.7
F1-Score	85.9

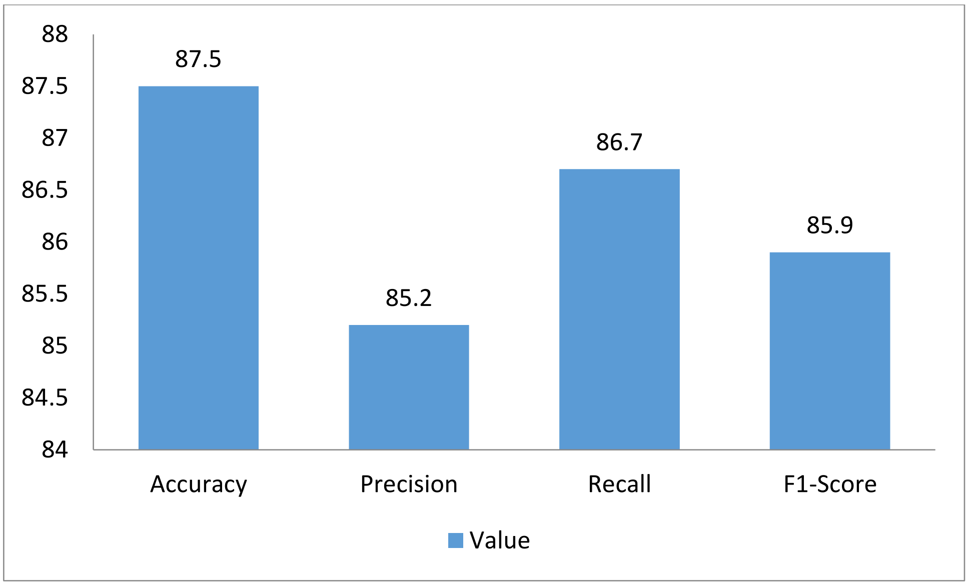

Fig 2 - Performance Table

ISSN: 3071-642X
PP: 01-12.

4.2.1. Accuracy (87.5%)

Accuracy is the percentage of all predictions (global correct rate) that are correct. An accuracy of 87.5% means that the model correctly predicted the market direction of most of the testing samples. The high value illustrates the reliability of stock market prediction when integrating financial sentiment analysis with technical indicators and employing LSTM network, which is more reliable than the traditional methods.

4.2.2. Precision (85.2%)

Precision is the percentage of positive predictions that are correct. The proposed model's accuracy reached 85.2 per cent, meaning most of the positive market movements predicted was correct. The accuracy is especially crucial when it comes to financial forecasting, as it aids in reducing the occurrence of false positive forecasts and helps investors make informed investment decisions, thereby limiting unrequired trading risks.

4.2.3. Recall (86.7%)

Recall is the percentage of the model's correct prediction of all actual positive market movements. The recall of 86.7% obtained indicates that the proposed system has been able to identify most of the positive market trends present in the dataset. A high recall value means that this prediction model is not likely to overlook important investment opportunities, which enhances its usefulness for real-life financial applications.

4.2.4. F1-Score (85.9%)

The F1 score is the harmonic average of precision and recall, and is used to measure the accuracy of the model's classifications. The proposed model attained an F1 score of 85.9%, which suggests the model has a fair balance between its ability to predict the correct direction of the market and the number of positive market changes it will detect. The result indicates that the model has a consistent performance across various evaluation measures and can be used for reliable financial sentiment analysis and stock market prediction.

4.3. Discussion

The experimental results illustrate the effectiveness of the proposed sentiment-infused stock market prediction model over the traditional models based on price data and technical indicators from historical data. The traditional approach to the analysis of market information, used in forecasting, mainly relies on numerical data like the price of stocks, the volume of trading, moving averages, and momentum indicators. These will help you see historical trends but will not help you if there is a sudden change in the market, due to investments emotions, breaking news, company announcements, or economic events. The proposed model integrates financial sentiment from news articles, social media platforms, and official company disclosures, providing more insights into market sentiment and investor expectations, which improves the accuracy and comprehensiveness of predictions. The accuracy of 87.5%, precision of 85.2%, recall of 86.7%, and F1-score of 85.9% suggest that the model is capable of accurately predicting market moves in the future with high reliability. The findings also show that positive sentiment tends to align with positive price movements, suggesting that positive financial information and investor sentiment drive more purchases into markets and boost confidence. On the other hand, negative sentiment is hard to tie to falling stock prices, as bad news, weak earnings, or adverse economic conditions can cause selling pressure and market volatility. These insights are consistent with the importance of investor sentiment on market dynamics. Additionally, LSTM networks have been shown to be effective for capturing long-term temporal dependencies in sequential financial data, which significantly contributes to the enhancement of the prediction performance. The major difference between LSTM and conventional machine learning algorithms is that it preserves significant information from past timesteps in its memory cells and discards the irrelevant information through the input, forget, and output gates. This ability allows the model to identify intricate patterns in historical market movements and sentiment indicators, leading to more accurate short-term market predictions. Combining textual sentiment analysis with technical market indicators can thus give a fuller picture of market dynamics than either technique can independently. In conclusion, the proposed hybrid strategy provides a strong, accurate, and practical approach to financial sentiment analysis and stock market prediction, which is valuable for investors, financial analysts, and decision-makers looking to obtain accurate market predictions.

5. CONCLUSION

The researchers developed a comprehensive framework that combines textual financial sentiment derived from financial news articles, financial statements, and social media platforms with historical stock market data and technical indicators to accurately forecast stock market trends and sentiments. The proposed technique integrates qualitative data on investor emotions with market quantitative indicators to offer a more comprehensive view of market behavior, as opposed to traditional prediction models that are mostly based on historical price changes and trading activity. Public perception, economic events, corporate disclosures, and investor confidence significantly shape the financial markets, making public opinion or sentiment analysis a crucial part of contemporary stock prediction systems. The study used preprocessing methods like tokenization, stop-word removal, stemming, and lemmatization, followed by the extraction of the features using TF-IDF and word embeddings. Technical indicators were incorporated into these features and then provided to a Long Short-Term Memory (LSTM) network that has capability to learn long-term dependencies in sequential financial data. The results of the experimental assessment showed that the proposed model had high predictive performance with accuracy of 87.5%, precision of 85.2%, recall of 86.7%, and F1 score of 85.9%. The results show that the addition of the financial sentiment index has a positive effect on the accuracy of prediction, as compared to the traditional models relying only on historical market data. The results also revealed that overall positive sentiment is linked with positive market movements and vice versa, negative sentiment is linked with negative market movements, thereby signifying the significance of investor sentiment in investment decisions. Moreover, the LSTM network successfully learned the temporal relationships between changes in sentiment and past stock prices thereby enhancing its short-term forecasting accuracy and robustness in different market scenarios. The framework suggested in this article shows that using textual and numerical data together can give better information on the trend of the market and can help investors, financial analysts, portfolio managers and automated trading systems to make smarter investment decisions. The proposed approach delivered good results, but some challenges still exist. Sentiment data can still have affecting factors such as availability, quality, noisy or misleading social media content and rapidly changing market circumstances. In conclusion, future studies should aim to explore how to embed sophisticated transformer-based text representations, like BERT, FinBERT and GPT-based financial language models, which capture more nuanced relationships in financial texts. Further studies could also focus on financial sentiment analysis in multiple languages to enhance global financial markets, integration of other data types like earnings call transcripts and macroeconomic reports, and building financial sentiment prediction systems that can process streaming data in real-time for high-frequency trading. In conclusion, the study highlights that financial sentiment analysis (WSIA) combined with technical market indicators and deep learning methods can be a valuable and effective tool for enhancing the accuracy of stock market predictions and aiding in intelligent financial decision-making in volatile markets.

REFERENCES

[1] Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35–65.
[2] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification Using Machine Learning Techniques. Proceedings of EMNLP, 79–86.
[3] Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Proceedings of ECML, 137–142.
[4] McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naïve Bayes Text Classification. AAAI Workshop on Learning for Text Categorization.
[5] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
[6] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.
[7] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT.
[8] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and Their Compositionality. Advances in Neural Information Processing Systems (NeurIPS).
[9] Vaswani, A., et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems (NeurIPS).
[10] Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance, 62(3), 1139–1168.
[11] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market. Journal of Computational Science, 2(1), 1–8.
[12] Si, J., Mukherjee, A., Liu, B., Li, Q., & Li, H. (2013). Exploiting Topic Based Twitter Sentiment for Stock Prediction. Proceedings of ACL.
[13] Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), 383–417.
[14] Fischer, T., & Krauss, C. (2018). Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. European Journal of Operational Research, 270(2), 654–669.
[15] Akita, R., Yoshihara, A., Matsubara, T., & Uehara, K. (2016). Deep Learning for Stock Prediction Using Numerical and Textual Information. Proceedings of the IEEE/ACIS International Conference on Computer and Information Science.