Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2024) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 10th International Conference on Computer Science and Computational Intelligence 2025 (ICCSCI 2025)

Deep Learning-based Stock Market Prediction Using Historical Price and Topic-Distributed News Sentiment

Rafi Hazel Tafara[a,*], Andrew[a], Yosepril Zhounggi[a], Muhammad Fikri Hasani[a], Ayu Maulina[a]

[a]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*

Abstract

Predicting stock prices remains a challenging task because of the volatile and non-linear nature of financial markets. This study proposes a hybrid deep learning method that combines sentiment and topic modeling of financial news with historical stock data to predict Apple Inc. stock prices. FinBERT is used to obtain sentiment scores and BERTopic is used to derive topic distributions and capture topic-aware sentiment context. These features are aligned on a daily basis and processed in a dual-branch neural architecture. One branch uses either LSTM, GRU, or Temporal CNN (TCNN) to handle historical price sequences, while the other is used to process topic-sentiment vectors. The results show that GRU performs slightly better than the other two models, with the highest R2 (0.9866), the lowest RMSE (2.81), and the lowest MAE (2.27). This study focuses only on Apple stock so it gives a more focused view on how financial news sentiment can affect the price prediction of a single stock. Limitations related to dataset size, computational resources, and real-time data availability are also discussed for future research.

Keywords: Stock Price; LSTM; GRU; TCNN; Sentiment Analysis; BERTopic; FinBERT; Apple Stock; Deep Learning; Financial New

1. Introduction

Stock market prediction has been a challenging problem that has fascinated researchers and investors for decades. Finding a method that can predict stock prices with significant accuracy is very helpful for investors in gaining the

* Corresponding author
E-mail address: rafi.tafara@binus.ac.id

1877-0509

optimal profit out of their investments by giving them information about the best time to buy and sell stocks. However, several obstacles remain that make predicting the future behavior of stock prices harder, such as the complexity of stock-related data and the volatility of prices. This is largely because stock prices are heavily influenced by a lot of information at the same time, such as the company's financial metrics [1], macroeconomic situation [2], and events with negative public sentiment—typically events that are geographically near the stock market's region [3]. Additionally, some researchers believed that the stock market itself is fundamentally unpredictable. Fama [4] suspected that the value of stock prices already reflects all publicly available information, implying that forecasting future prices is impossible with only public information; a hypothesis commonly known as the Efficient Market Hypothesis (EMH). Though this hypothesis is still debatable for many years as more research was done using new forecasting methods with mixed results, which lead some researchers argue that the stock market can be predicted to some extent [5].

Conventional methods on stock market prediction include fundamental analysis, technical analysis, and a combination of both. Advancement in computing technology and Artificial Intelligence (AI) has driven the utilization of Machine Learning (ML) techniques into stock market prediction by researchers and investors alike. Forecasting stock prices with ML usually revolves around finding patterns in the same data used in fundamental or technical analysis and extrapolating them. One of the earliest ML-based stock forecasting application was published in 1976 that uses a perceptron-like pattern recognition system [6] to determine the optimal investment decision—either "buy" or "sell". In the following decades, other, more rigorous ML models were used in most studies and practices. Popular models for stock prediction are Linear Regression (LR), Support Vector Machine (SVM), and Artificial Neural Network (ANN) [7]. Among these, ANN is more commonly used among all research papers [8] for its effectiveness in detecting nonlinear patterns compared to other technical methods [7]. While there exist several variants of ANN such as Recurrent Neural Network (RNN) [9] and Convolutional Neural Network (CNN) [10], Long Short-Term Memory (LSTM) [11] stands out as an improvement of RNN that can effectively process and predict data from long sequences.

Studies [12][13] show that models which incorporate more data sources—quantitative and qualitative—tend to yield more accurate predictions compared to models with less data sources. While traditional fundamental and technical analysis made heavy use of quantitative information, Li *et al.* [14] believed that qualitative information can complement quantitative data and improve the prediction model by also considering additional data, such as sentiment, hidden in between the texts from sources like news and social media. Some recent studies have combined sentiments from news or social media with historical stock prices in their model [15–17]. However, according to our literature review, only a handful of research papers use topic-like grouping on their textual data. We believe that some news topics have a greater influence on stock prices than other topics. Thus, capturing the magnitude of each topic's influence may improve the prediction model's accuracy.

To address that issue, in this paper, we propose an architecture that combines historical stock prices with a hybrid of financial news topic modeling and sentiment analysis to predict future stock prices. This is done by computing the topic model from the news corpus which returns the most probable topic of each document. Sentiment is then analyzed for each document, and the distribution of sentiments across every existing topic, together with historical stock prices, is trained using RNN-based models like LSTM.

2. Related Works

There are multiple approaches for predicting stock market prices that have been done in previous studies. Various methods have been developed to increase the prediction accuracy, ranging from fundamental analysis, technical analysis, and a combination of both to the implementation of AI and ML. Some of the previous studies use historical stock prices to increase the accuracy of the prediction model, such as [18]. Some other studies use news analysis sentiment to improve the accuracy of the prediction model, such as [19]. There are also studies that combine both historical data and news sentiment analysis to increase the accuracy of the model, such as [15–17].

2.1. Studies that are based on historical stock prices

Most of the conventional studies on stock market prediction use historical stock price data to find trends and patterns to predict future price changes. Several machine learning and deep learning methods, such as Support Vector Machines (SVM), Linear Regression (LR), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM), have been used in these studies [7]. P. Gupta. [18] used historical data from Yahoo Finance to analyze and predict stock prices using the Long Short-Term Memory (LSTM). TensorFlow or Keras was used to create the LSTM model, which predict closing prices, trading volumes, and daily returns. The study finds that Yahoo Finance, LSTM, and analytical tools like Pandas and Matplotlib will give investors useful information when making decisions.

2.2. Studies that are based on news sentiment analysis

K. Joshi. [19] explored the relationship between financial news sentiment and stock trends. The study uses several approaches like Naïve Bayes, Random Forest, and Support Vector Machine (SVM). The model's performance was evaluated using Apple Inc. stock data from 2013 to 2016. The study finds that Random Forest achieved the highest accuracy (88-92%), SVM (86%), and Naïve Bayes (83%). The study showed that sentiment analysis could significantly enhance stock trend prediction, outperforming a random baseline by 30%.

2.3. Studies that are based on both news sentiment analysis and historical prices

A. E. Khedr. [15] explains how to use data mining techniques and news sentiment analysis to predict stock market behavior. The suggested model uses a dataset from three NASDAQ-listed companies and combines the Naïve Bayes algorithm for sentiment analysis with K-Nearest Neighbor (KNN) for stock price prediction. The study's results indicate that combining news sentiment analysis with historical stock prices improves prediction accuracy by up to 89.80%, demonstrating that financial news significantly influences stock price movements.

X. Li. [16] uses a Long Short-Term Memory (LSTM) network to combine technical indicators and news sentiment to create a stock prediction system. The Loughran–McDonald Financial Dictionary is the most successful of the four sentiment dictionaries compared in this study, which uses data from the Hong Kong Stock Exchange spanning more than five years. The significance of integrating market information in predictive analysis is highlighted by experimental results that demonstrate that the LSTM-based model incorporating both stock prices and news improves prediction accuracy by up to 120% when compared to models using only one type of data.

K. Fu. [17] suggests the BERT-LLA model, which combines BERT-based sentiment analysis with stock price prediction using LSTM enhanced with an attention mechanism. His model takes into account sentiment from a variety of sources, such as stock technical indicators and investor and financial news comments. To increase the accuracy of stock price predictions, its novel strategy also uses sector heat and semantic similarity to choose related industries. According to experimental findings, this model outperforms earlier approaches in capturing how market sentiment affects changes in stock prices.

2.4. Studies that are based on both news sentiment analysis and topic modeling

There are several approaches to stock movement prediction that integrate both financial data and textual analysis. Past studies have developed prediction models using historical data, sentiment analysis, topic modeling, and combinations of these methods. One prominent line of research P. Hajek. And A. Brushka. [20] uses sentiment analysis from financial news to improve forecasting accuracy. Other studies employ topic modeling to uncover hidden themes in financial text that may influence investor behavior. Recent works have also explored the integration of sentiment and topic detection using advanced machine learning models, such as deep neural networks, to further enhance predictive power.

2.5. Studies that are based on deep learning for stock prediction

A. Thakkar. [21] give a thorough analysis of the application of Deep Neural Networks (DNNs) in stock market prediction. The study explains several of DNN models like CNN, RNN, LSTM, GRU, ESN, DBN, and RBM. These models are helpful because they can automatically identify significant patterns in stock data. The study also highlights some common issues like overfitting, difficulty understanding how the models work, and the significance of feature selection. This study also did an experiment with nine deep learning models using the S&P 500 dataset to test how well these models predict stock trends and how the quantity of features can affect the result.

2.6. Studies that are based on Indonesian stock data

Hotasi. And D. P. Satya [22] adapted the same two stage framework to the Indonesian stock market. Their study applied a Naïve Bayes classifier to financial news related to stocks listed on the IDX (BBCA, BUMI, ASII) to determine sentiment polarity. The sentiment scores were then combined with historical stock data and used as input for a KNN classifier to predict future stock trends. The model achieved 91.9% - 97.3% accuracy in sentiment classification and up to 90% in stock trend prediction during the testing phase. This research represents one of the first applications of such hybrid modeling to Indonesian financial data, bridging the gap between text mining and traditional financial analysis.

3. Methodology

The stock price prediction architecture proposed by us can be seen in Figures below.

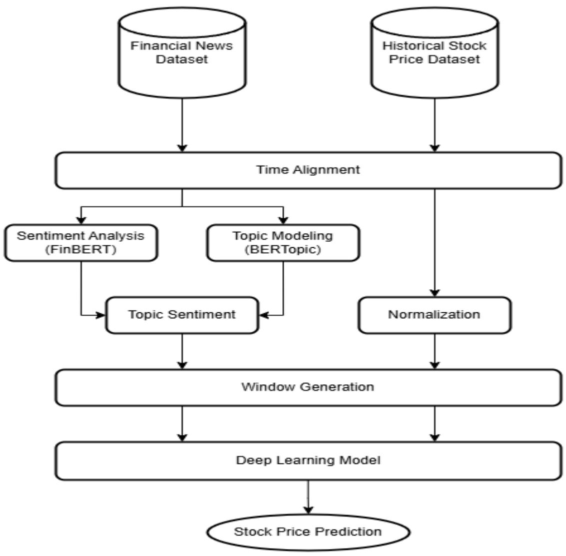

Fig. 1. General architecture.

The diagram in Fig. 1 above illustrates the general architecture of our proposed model for stock price prediction. Although the diagram captures the overall structure, including topic sentiment and historical price branches, some implementation-specific details have been simplified or left out for clarity.

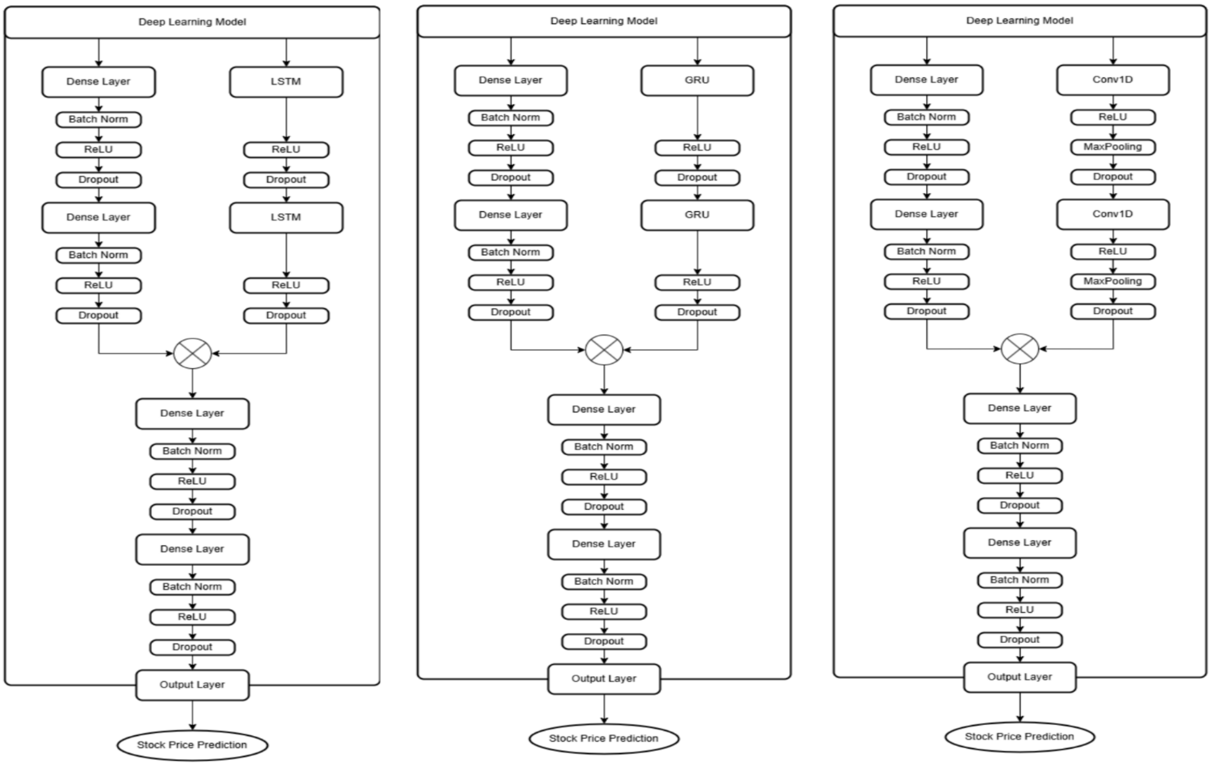

Fig. 2. Deep Learning using LSTM, GRU, and TCNN.

3.1. Time Alignment

Every data must be aligned according to their day. This is trivial for historical price data P_t. For news data, all documents are aggregated according to their date of publication. Let D be the set of all news documents and δ_t be the set of all news documents published on day t. Mathematically, any given news document D_d is a part of δ_t if D_d is published on day t.

3.2. News Sentiments Analysis

Sentiment analysis is one of the tasks of Natural Language Processing (NLP) that classifies the emotional tone expressed in a text. Commonly, it determines whether the tone of a given text is positive or negative. Stock prices are known to be influenced by news sentiments to some degree, particularly financial news as seen in other related papers such as [19]. The architecture proposed in this paper uses FinBERT [23] to analyze the sentiment of news text from the dataset. FinBERT is a pre-trained language model fine-tuned from Google's BERT language model [24] with Financial PhraseBank [25] dataset which contains pairs of financial news sentences and its sentiment label. In theory, this makes FinBERT more optimized at handling sentiment analysis tasks on our financial news dataset.

For each news document D_d, FinBERT outputs a softmax-activated probability distribution of three sentiment labels: positive, negative, and neutral. Each probability value lies inside the range $[0, 1]$. Let the probability value of a positive and negative sentiment label be $\text{FinBERT}_{+}(D_d)$ and $\text{FinBERT}_{-}(D_d)$ respectively. The sentiment score S for each news is calculated as follows:

$$S(D_d) = \text{FinBERT}_{+}(D_d) - \text{FinBERT}_{-}(D_d) \quad (1)$$

This results in a variable that ranges between $[-1, 1]$. A sentiment score closer to 1 indicates a stronger positive sentiment, while a sentiment score closer to –1 indicates a stronger negative sentiment. This also implies that values close to 0 indicate a neutral sentiment.

3.3. News Topic Modeling

Topic modeling (TM) is another branch of NLP that focuses on clustering text documents in a corpus into several topics or categories by comparing each document's similarity with each other. Instead of simply aggregating the sentiment score out of every document as often seen in other related papers, this step clusters all the news documents based on similarity which effectively distributes every news document and its sentiment score into each topic. This way, an array of topics containing its aggregate topic sentiment score can be used as additional features for the deep learning model. The intention here is to enhance the overall model's accuracy by optimizing the sentiment score's relevance to a stock as the sentiment value on certain topics could have more influence than others. By implementing this step, influence from certain topics can be more precisely attributed.

The architecture proposed here employs BERTopic [26] to create the topic distribution of the news dataset. BERTopic itself is a modular TM pipeline that primarily consists of three steps. First, it computes embeddings using a pre-trained language model. Then, after reducing the dimensionality of the embeddings, it clusters documents that are semantically similar, forming a topic. Finally, a topic representation is extracted from every topic using a custom class-based TF-IDF.

BERTopic clusters the news documents into K different clusters/topics, where the exact amount is subject to change with hyperparameter tuning. For each news document D_d, BERTopic outputs a K-length array of topic probability distribution with values ranging between $[0, 1]$. This information is useful especially for news documents that are not necessarily discussing only a single topic (i.e. contains a mixture of topics). By multiplying this array with $S(D_d)$, the sentiment score is effectively distributed according to its topic distribution—dominant topics receive a larger portion of its sentiment score, and vice versa.

3.4. Topic Sentiment

The following equation calculates the topic sentiment array TS_t, an array that combines the news sentiment score and topic distribution at day t. Let $\delta_{t,i}$ be an arbitrary news document in δ_t. For each news document $\delta_{t,i}$ in δ_t, let $\text{BERTopic}(\delta_{t,i})$ be the topic probability distribution of $\delta_{t,i}$. Each element of TS_t is defined as the average of the product between the topic probability and sentiment score defined in equation (1) out of all news documents in δ_t.

$$TS_t = \frac{\sum_i \left(S(\delta_{t,i}) \cdot \text{BERTopic}(\delta_{t,i})\right)}{\sum_i \text{BERTopic}(\delta_{t,i})} \quad (2)$$

3.5. Window Generator

Training an RNN-like model requires splitting the input data by their day and creating a "sliding window". Let Z_t be a $(K + 5)$-length array of features that correspond to day t. Z_t consists of two main parts: P_t and TS_t. A 360-length window W_i is defined in equation (3). Thus, for all valid values of $i \in [1, T - 359]$ where T represents the total number of days in the combined dataset, W_i is used as the training data for the deep learning model.

$$W_i = [Z_i, \dots, Z_{i+359}] \quad (3)$$

3.6. Data Normalization

Numerical data from the historical price dataset must be normalized first to prevent disproportional contribution in the training process from features with large values. This architecture implements Z-score normalization which scales the data into a normal distribution. Additionally, the mean μ and standard distribution σ is calculated for each window so that the scaled data can adapt to new values (i.e. new highs/lows). For each feature j in P_t, given the mean $\mu_{W_i,j}$ and standard deviation $\sigma_{W_i,j}$ in window W_i, the normalized value $P'_{t,j}$ is calculated as follows:

$$P'_{t,j} = \frac{P_{t,j} - \mu_{W_i,j}}{\sigma_{W_i,j}} \tag{4}$$

3.7. Deep Learning model pipeline

The deep learning network consists of two branches which will later merge together. The first branch processes the topic sentiment features from the news dataset, while the second branch processes the historical price data. The layers used in each branch are detailed below.

3.7.1. Topic Sentiment Branch

This branch contains two dense layers, each equipped with a batch normalization and a dropout layer to improve learning performance and reduce the risk of overfitting.

3.7.2. Historical Price Branch

This branch contains two LSTM/GRU/TCNN layers stacked together—each equipped with a dropout layer—to capture any sequential patterns in the data. The pipeline will be tested several times, each swapping this layer with either LSTM, GRU, or TCNN. The result will be compared to find out which network type performs better.

3.7.3. Merging Branch

The output vectors of the two previous branches are concatenated together and fed into two dense layers—each also equipped with a batch normalization and a dropout layer—and an output layer with a linear activation.

Computing the loss value is done using the next day's closing price as the target variable. The evaluation is done using loss functions such as MAE, RMSE, and R2 score. The deep learning network will be optimized using Adam optimizer. Any hyperparameter tuning done will be detailed in the next section.

4. Result and Discussion

The experiments use Apple stock price data from Yahoo Finance and news data from Kaggle. Results from the experiments are shown in this section. LSTM, GRU, and TCNN prediction models were used to analyze data on Apple stock prices that was scraped using the beautifulsoap library and collected between January 1, 2020, and April 30, 2025. The models are trained using a laptop with AMD Ryzen 7 6800, NVDIA RTX 3060 (6GB VRAM), and 16GB RAM. Data from the dataset includes date, open, high, low, close, and volume. Data from financial news was used in addition to the dataset of stock prices. Table dan figures below explains the different result of the selected models.

Table 1. Evaluation for the selected models.

Model	MAE	RMSE	R2
LSTM	2.45	3.10	0.9838
GRU	2.27	2.81	0.9866
TCNN	3.48	4.95	0.9587

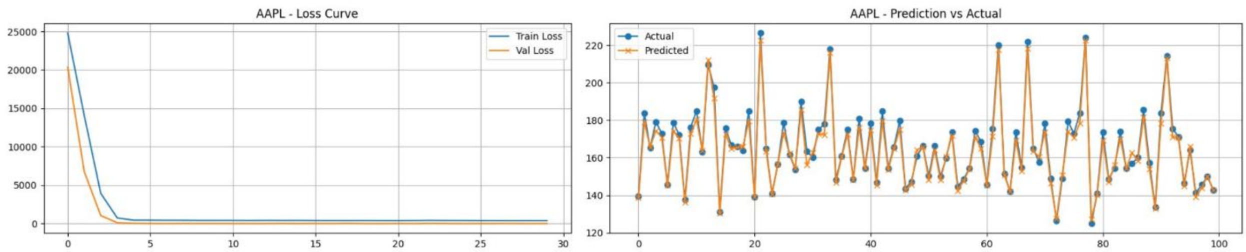

Fig. 3. LSTM results.

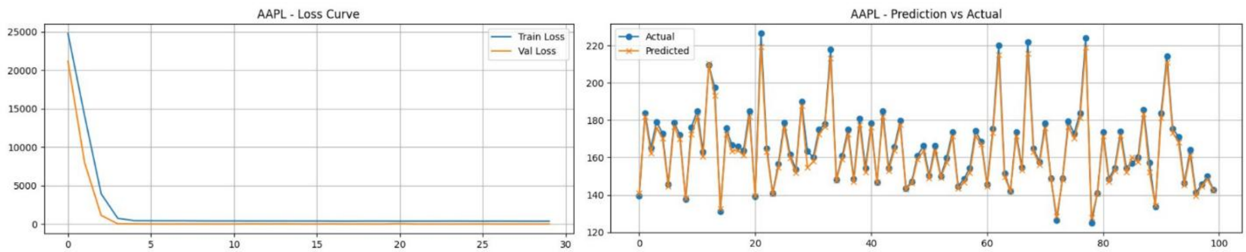

Fig. 4. GRU results.

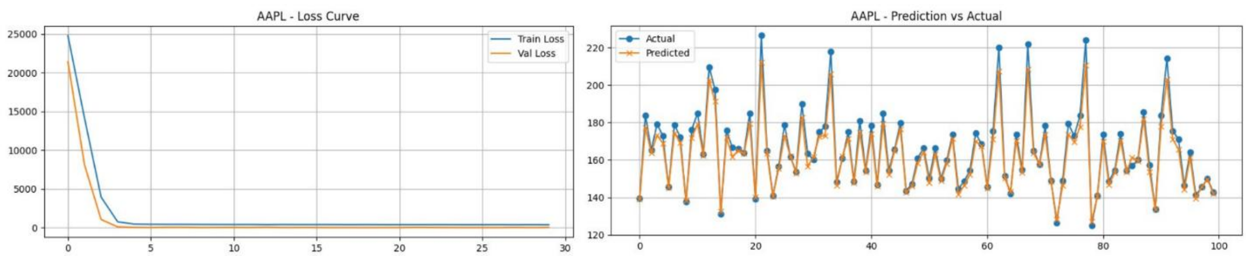

Fig. 5. TCNN results.

4.1. Loss Curve and Prediction Analysis

The training and validation loss curves (Figure 3-5) for all models show stable convergence after approximately 5 epochs, indicating that the models are not overfitting. The plots of predictions and actuals also demonstrate that while TCNN exhibits greater variance and more notable deviations, GRU and LSTM can closely follow actual price movements. The conclusion that recurrent models (GRU, LSTM) are better suited for time-dependent financial forecasting tasks is further supported by these visualizations, particularly when combined with sentiment-based features.

4.2. Discussion of limitations

There are several limitations in this study that may have affected the model's performance and generalizability. First, the dataset was limited to Apple Inc. (AAPL) stock and news articles sourced only from Apple News and CNN, which might have excluded important sentiment signals from other significant financial sources like Nasdaq or Bloomberg. Second, the 360-day sliding window restricted training efficiency on standard hardware with limited RAM due to its high memory requirements, even though it was useful for capturing long-term temporal patterns. Additionally, website restrictions and anti-bot protections made it difficult to scrape news in real-time, forcing reliance on static pre-collected datasets that may not fully capture dynamic market sentiment. Finally, the topic modeling process used sentence embeddings that were generated using a general-purpose model (MiniLM), which could be improved with domain-specific embeddings or fine-tuned transformer models for financial texts. By integrating more varied data sources, scalable infrastructure, real-time data pipelines, and domain-specific language models, future research could overcome these constraints.

5. Conclusion

The experiments showed that all three models could learn the correlation between topic-distributed sentiment signals and stock prices. Standard regression metrics (MAE, RMSE, and R2) were used to assess each model. The GRU-based and LSTM-based methods tended to yield somewhat more stable results in constrained hardware settings, while TCNN provided a computationally lighter option, even though the models displayed consistent performance trends. The results show that GRU slightly outperforming LSTM and TCNN in all metrics, with the highest R2 (0.9866), the lowest RMSE (2.81) b and MAE (2.27). For future work, the model could be enhanced by incorporating more diverse and up-to-date news sources, experimenting with more robust embedding techniques, and implementing a more efficient architecture to support real-time forecasting scenarios. With these improvements, the model may offer better generalization and adaptability for use in real-world financial forecasting applications.

6. Acknowledgements

6.1. Data Availability Statement

The datasets used in this study are available from a publicly accessible source. Historical price data were obtained from Yahoo Finance at https://finance.yahoo.com/quote/AAPL/history/. Apple Stock (AAPL) historical financial news dataset were obtained from Fujoos's dataset on https://www.kaggle.com/datasets/frankossai/apple-stock-aapl-historical-financial-news-data. CNN news articles were obtained from Hadas Unger's dataset on https://www.kaggle.com/datasets/hadasu92/cnn-articles-after-basic-cleaning.

6.2. Credit Authorship Contribution Statement

Rafi Hazel Tafara: Conceptualization, Modeling, Dataset Gathering, Writing – Review & Editing; **Andrew:** Conceptualization, Dataset Gathering, Training model, Experiment, Writing – Review & Editing; **Yosepril Zhounggi:** Conceptualization, Dataset Gathering, Writing – Review & Editing; **Muhammad Fikri Hasani:** Supervision, Writing – Review & Editing; **Ayu Maulina:** Supervision, Writing – Review & Editing;

References

[1] Chalise DR. Does Financial Information Influence The Behavior of Stock Price? Evidences From Nepali Commercial Banks. Saptagandaki Journal 2022:146–57. https://doi.org/10.3126/sj.v13i1.54952.

[2] Hardouvelis GA. Macroeconomic information and stock prices. J Econ Bus 1987;39:131–40. https://doi.org/10.1016/0148-6195(87)90012-9.

[3] Ichev R, Marinč M. Stock prices and geographic proximity of information: Evidence from the Ebola outbreak. International Review of Financial Analysis 2018;56:153–66.

https://doi.org/10.1016/j.irfa.2017.12.004.
[4] Fama EF. Efficient Capital Markets: A Review of Theory and Empirical Work. J Finance 1970;25:383. https://doi.org/10.2307/2325486.
[5] Chen C, Dongxing W, Chunyan H, Xiaojie Y. Exploiting Social Media for Stock Market Prediction with Factorization Machine. 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE; 2014, p. 142–9. https://doi.org/10.1109/WI-IAT.2014.91.
[6] Felsen J. A man-machine investment decision system. Int J Man Mach Stud 1976;8:169–93. https://doi.org/10.1016/S0020-7373(76)80042-9.
[7] Shi Y. Research on the Stock Price Prediction Using Machine Learning. Advances in Economics, Management and Political Sciences 2023;22:174–9. https://doi.org/10.54254/2754-1169/22/20230307.
[8] Nti IK, Adekoya AF, Weyori BA. A systematic review of fundamental and technical analysis of stock market predictions. Artif Intell Rev 2020;53:3007–57. https://doi.org/10.1007/s10462-019-09754-z.
[9] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences 1982;79:2554–8. https://doi.org/10.1073/pnas.79.8.2554.
[10] LeCun Y, Haffner P, Bottou L, Bengio Y. Object Recognition with Gradient-Based Learning, 1999, p. 319–45. https://doi.org/10.1007/3-540-46805-6_19.
[11] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput 1997;9:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.
[12] Zhang X, Zhang Y, Wang S, Yao Y, Fang B, Yu PS. Improving stock market prediction via heterogeneous information fusion. Knowl Based Syst 2018;143:236–47. https://doi.org/10.1016/j.knosys.2017.12.025.
[13] Zhang X, Qu S, Huang J, Fang B, Yu P. Stock Market Prediction via Multi-Source Multiple Instance Learning. IEEE Access 2018;6:50720–8. https://doi.org/10.1109/ACCESS.2018.2869735.
[14] Li Q, Jiang L, Li P, Chen H. Tensor-Based Learning for Predicting Stock Movements. Proceedings of the AAAI Conference on Artificial Intelligence 2015;29. https://doi.org/10.1609/aaai.v29i1.9452.
[15] Khedr AE, S.E.Salama, Yaseen N. Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis. International Journal of Intelligent Systems and Applications 2017;9:22–30. https://doi.org/10.5815/ijisa.2017.07.03.
[16] Li X, Wu P, Wang W. Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. Inf Process Manag 2020;57. https://doi.org/10.1016/j.ipm.2020.102212.
[17] Fu K, Zhang Y. Incorporating Multi-Source Market Sentiment and Price Data for Stock Price Prediction. Mathematics 2024;12:1572. https://doi.org/10.3390/math12101572.
[18] Gupta P, Malik S, Apoorb K, Sameer SM, Vardhan V, Ragam P. Stock Market Analysis using Long Short-Term Model. ICST Transactions on Scalable Information Systems 2023. https://doi.org/10.4108/eetsis.4446.
[19] Joshi K, H. N B, Rao J. Stock Trend Prediction Using News Sentiment Analysis. International Journal of Computer Science and Information Technology 2016;8:67–76. https://doi.org/10.5121/ijcsit.2016.8306.
[20] Hajek P, Barushka A. Integrating Sentiment Analysis and Topic Detection in Financial News for Stock Movement Prediction. Proceedings of the 2nd International Conference on Business and Information Management, New York, NY, USA: ACM; 2018, p. 158–62. https://doi.org/10.1145/3278252.3278267.
[21] Thakkar A, Chaudhari K. A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. Expert Syst Appl 2021;177. https://doi.org/10.1016/j.eswa.2021.114800.
[22] Hotasi A, Satya DP. Predicting Stock Market Trends with News Sentiment Analysis and Historical Stock Market Data Analysis Using Data Mining Techniques. 2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA), IEEE; 2023, p. 1–6. https://doi.org/10.1109/ICAICTA59291.2023.10390037.
[23] Araci D. FinBERT: Financial Sentiment Analysis with Pre-trained Language Models 2019. https://doi.org/10.48550/arXiv.1908.10063.
[24] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding 2018. https://doi.org/10.48550/arXiv.1810.04805.
[25] Malo P, Sinha A, Korhonen P, Wallenius J, Takala P. Good debt or bad debt: Detecting semantic orientations in economic texts. J Assoc Inf Sci Technol 2014;65:782–96. https://doi.org/10.1002/asi.23062.
[26] Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure 2022. https://doi.org/10.48550/arXiv.2203.05794.