On the Impact of News for Reliable Stock Market Predictions: An LSTM-based Ensemble using FinBERT Word-Embeddings

MOHSEN A. HASSAN
Information System, Faculty of Computers & Artificial Intelligence, Helwan University, Cairo, EGYPT

ALIAA AA YOUSSIF
Computer Science, Arab Academy for Science, Technology and Maritime Transportation (AASTMT), Cairo, EGYPT

OSAMA IMAM
Information System, Faculty of Computers & Artificial Intelligence, Helwan University, Cairo, EGYPT

AMR S. GHONEIM
Computer Science, Faculty of Computers & Artificial Intelligence, Helwan University, Cairo, EGYPT

Abstract: - Stock market (SM) prediction methods can be divided into two categories based on the number of information sources used: single-source methods and dual-source approaches. To estimate the price of a stock, single-source approaches rely solely on numerical data. The Efficient Market Hypothesis (EMH), [1]. States that the stock price will represent all important information. Different sources of information might complement one another and influence the stock price. Machine learning and deep learning techniques have long been used to anticipate stock market movements, [2], [3]. The researcher gathered the dataset, [4], [5], [6], [7]. The dataset contains the date of the reading, the opening price, the high and low value of the stock, news about the stock, and the volume. The researcher uses a variety of machine Learning and deep learning approaches to compare performance and prediction error rates, in addition, the researcher also compared the effect of adding the news text as a feature and as a label model. and using a dedicated model for news sentiment analysis by applying the FinBERT word embedding and using them to construct a Long Short-Term Memory (LSTM). From our observation, it is evident that Deep learning-based models performed better than their Machine learning counterparts. The author shows that information extracted from news sources is better at predicting rather than its direction of price movement. And the best-performing model without news is the LSTM with an RMSE of 0.0259 while the best-performing model with news is the LSTM with a stand-alone and LSTM model for news yields RMSE of 0.0220.

Key-Words: - Dual-Sources Stock Market Prediction, BERT Word-Embedding, Long Short-Term Memory (LSTM), Stock Price Prediction, Wealth Management.

Received: October 13, 2021. Revised: September 6, 2022. Accepted: October 9, 2022. Published: November 7, 2022.

1 Introduction

Wealth Management (WM) is an asset planning and structuring discipline that aids in wealth creation, preservation, and protection, [8]. One of the methods proposed by wealth management gurus to build one's wealth is to invest in the stock market. The stock market is a significant contributor to a country's economic development, [9]. It's an opportunity for investors to purchase a brand-new stock and either become a stockholder who receives a shareholder bonus, or a stock trader who trades stock on the stock exchange. A stock trader could generate more revenue if they correctly identified and predicted stock price patterns. The stock market is by its very nature quite volatile. Daily news developments, such as shifting political situations, a firm's performance, and other unanticipated events, have an immediate favorable or negative impact on stock values, [10]. As a result, accurately predicting stock prices and directions is difficult; investors must look for long-term trends. Traditional analytical methods are widely used in the fields of economics and finance, [11], [12], [13], [14]. and they rely on fundamental and technical analysis. The fundamental analysis technique, [15], 16], [17]. investigates external elements that affect the stock's intrinsic value, including interest rates, currency rates, inflation, industrial policy, listed company finance, international relations, and other economic and political aspects. On the other hand, the technical analysis method primarily focuses on the direction of stock price, trading volume, and investors' psychological expectations. It focuses on using Kline charts and other tools to analyze the stock index trajectory of individual stocks or the entire market. The number of information sources employed in stock market prediction methods can be classified into two categories: single-source techniques and dual-source approaches, [18], [19], [20]. Single-source techniques depend entirely on numerical data to predict the price of a stock. According to the Efficient Market Hypothesis (EMH), [21], the stock price will reflect all relevant information. The stock price may be influenced by different sources of information that complement one another. Thus, the dual-source approaches focus on developing appropriate news representations while capturing the data's temporal relationship, [22], [23], [24]. In recent years, however, both the rate of publication and the number of daily news providers have risen dramatically, considerably outstripping investors' ability to sift through massive amounts of data. As a result, an automated decision-making system is essential to analyze and forecast future stock movements. One of the most challenging tasks for both traders and academics /researchers is stock market forecasting, [25]. Because of its enormous earning potential, the stock market has always attracted a large number of investors. Researchers believe stock market prediction is challenging due to the difficulty in obtaining nonlinear and non-stationary variance in data, [26]. Thus, Stock market forecasting has long relied on machine learning and deep learning approaches, [27], [28]. The development of Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM), [29]. Attention Mechanisms [30], particularly Self-Attention and Transformers, [31], are examples of recent developments in deep learning. These methodologies have significantly increased the accuracy of word-based tasks such as sentiment analysis prediction, [32]. As a result, this paper proposes a successful model for sentiment analysis of stock market-related news that uses BERT (Bidirectional Encoder Representations from Transformers), [33]. word-embedding and LSTMs.

After stating the motivations for analyzing stock market-related news, the utilization of LSTMs, and highlighting the difference between single-source and dual-source techniques, the remaining part of the paper is organized as follows. Selected relevant approaches for dual-source stock market prediction (mostly using deep learning approaches) are reviewed in Section 2. Section 3 presents the proposed methodology, including the used word embedding and the planned pipeline. In Section 4, a description of the constructed dataset is given. The results are presented and discussed in Sections 5 and 6, respectively. Finally, the conclusion and further work are found in Section 7

2 Literature Review

A Numerical-based Attention (NBA) approach for dual-source stock market prediction was proposed by, [34]. First, they proposed an attention-based stock price prediction strategy that effectively harnesses the complementarity of news and numerical data. The stock trend information hidden in the news is captured by the crucial distribution of numerical data. As a result, the information is encoded to make numerical data selection easier. Their approach effectively filters out noise while boosting the usefulness of news trend information. Then, three datasets were created using a news corpus and numerical data from two sources to evaluate the NBA model. And with the advancement of text mining techniques, [35]. proposed a modern autoregressive neural network architecture that incorporated sentiment predictors. They suggested that using predictors based on counts of news articles/stories and Twitter posts will considerably improve the accuracy of stock price predictions. Also, in, [36]. projected the stock price movements by exploring a stock price prediction based on news sentiment analysis. In addition, the authors proposed using sentiment analysis to rate articles using single combined strings and a positive, negative, or neutral rating string. The performance of the sentiment analysis is incorporated into any machine learning models that predict the stock market. Instead of utilizing complete news articles, [37]. concentrated on the economic news headlines. They employed several approaches to analyzing the sentiment of the headlines. They employed BERT as a baseline and then used other tools (namely, VADER, Text Blob, and a Recurrent Neural Network) to compare the sentiment analysis findings to stock changes over the same period. They concluded that both BERT and the RNN could assess emotional values without neutral parts far more accurately than the other two tools. By comparing these results to the behavior of stock market prices over the same periods via sentiment analysis of economic news headlines, they were able to determine the timings of the changes in stock values. Because the initial weight of the random selection issue is easily prone to erroneous predictions, traditional neural network algorithms may incorrectly predict the stock market when examining the influence of market factors on stock prices. Based on the development of word vectors in deep learning, [38]. presented the concept of a stock vector. Thus, rather than a single index or single stock index, the input is multi-stock high-dimensional historical data. Pang et al recommended using a Deep Long Short-Term Memory (LSTM) Neural Network with an embedded layer and an LSTM Neural Network with an automatic encoder to predict the stock market, [36]. investigated the first impact of COVID-19 sentiment on the United States (US) stock market using big data such as the Daily News Sentiment Index (DNSI) and Google Trends data on coronavirus-related searches. The goal was to examine if there was a correlation between COVID-19 sentiment and 11 distinct stock market sector indexes in the US over a specified period.

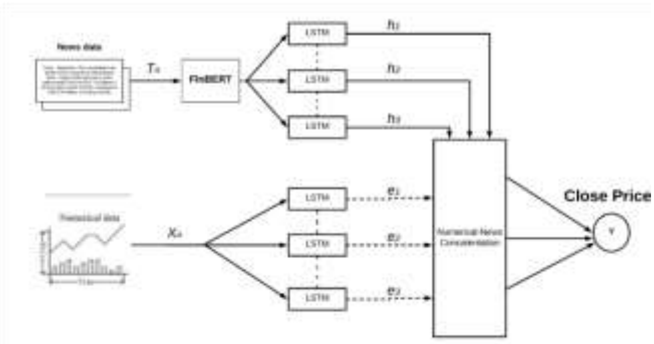

Fig. 1: Proposed Pipeline.

Any positive or negative public sentiment regarding stock market crises could have a cascade effect on stock market decision-making. The data showed how the COVID-19 sentiment had distinct effects on the different industries, and they were then categorized into different correlated groups, [39]. propose a multi-source multiple instance model capable of combining events, sentiments, and quantitative data into a comprehensive framework. It is difficult to work with qualitative data since it is typically unstructured, making it difficult to extract important signals from it. Because both events and sentiments can influence market fluctuations, it is reasonable to examine how to efficiently combine them to generate a better prediction. They offer various Instance Learning (MIL) model extension that effectively integrates multiple sources to create more accurate predictions. To distribute attention on the most effective day, [40]. used Bidirectional-LSTM followed by a self-attention mechanism. They evaluated their model on the Standard Poor's 500 index and individual stock prices, demonstrating that their baseline is competing with the existing state-of-the-art model, [41]. Enable a novel strategy for simplifying noisy-filled financial temporal series by sequence reconstruction using motifs (frequent patterns), and then use a convolutional neural network to capture time series spatial structure. The proposed framework outperforms established methods that use frequency trading patterns, improving accuracy by 4% to 7%, [42]. Propose a novel embedding technique that treats the news node as a set of features, each of which is produced using a sub-node model. The news is represented as a vectorial concatenation of features. The used pipeline outperforms the previous state-of-the-art.

3 Methodology

The researchers use a variety of machine Learning approaches (K-Nearest Neighbors - KNN, Decision Tree, Random Forest Regressor, Light Gradient Boosting Machine, Gradient Boosting Regressor, ADABOOST Regressor, Extra Trees Regressor) and deep learning approaches (Long Short-Term Memory - LSTM, Bidirectional Long Short-Term Memory - BI-LSTM, Gated Recurrent Unit - GRU, Bidirectional Gated Recurrent Unit BI-GRU, Convolutional Neural Network - Long Short-Term Memory - CNN-LSTM, ATTENTION-LSTM, and ATTENTION- BI-LSTM and ATTENTION-GRU and ATTENTION-BI-GRU) to compare performance and prediction error rates, and investigating how modern machine and deep learning techniques can be utilized for stock market prediction while testing on four datasets.

In addition, the researcher compared the performance of models that include news and models that do not include news in their training datasets.

The researcher also compared the effect of adding the news text as a feature and as a label model. and using a dedicated model for news sentiment analysis From our observation, using a dedicated model for news sentiment analysis proved to be more effective

than adding the news in one model as a labeled value.
The researcher proposed the model's pipeline commences by applying the FinBERT word embedding to the news data (described in Section 4) and using them to construct (i.e., train) a Long Short-Term Memory (LSTM).

3.1 Model Architecture

In this section, the proposed model is presented (as shown in Figure 1). The model's pipeline commences by applying the FinBERT word embedding [43] to the news data (described in Section 4), and using them to construct (i.e., train) a Long Short-Term Memory (LSTM) [29]. Simultaneously, another LSTM is trained using the numerical data. Finally, both models are then integrated, thus allowing them to utilize all features extracted by both models (from the "Numerical + News" Data) to predict the closing prices, consequently yielding reduced root mean square errors (RMSEs).

3.2 FinBERT Embedding

Intended for the embedding, FinBERT, [43]. which is a language model based on BERT, [33]. is utilized. FinBERT has been developed and employed to tackle natural language processing (NLP) tasks in the financial domain, but it has not been applied in dual-source stock market predictions that incorporate news data. BERT has created a stir in the machine learning field by delivering cutting-edge results in a wide range of NLP tasks, including Question Answering, [44]. Natural Language Inference, among others. BERT's main technical breakthrough is the use of a Transformer's bidirectional training for language modeling (the transformer is a popular attention model, [31]. This differs from earlier research, [45], which focused on a text sequence from left-to-right, or a combination of left-to-right and right-to-left training. The findings within the literature suggest that bidirectionally trained language models can have a better understanding of language context and flow than single-direction language models. Transformers are an attention mechanism that learns contextual relationships between words or sub-words in a text, and BERT makes use of them. The transformer as shown in Figure 2 – in its basic arrangement – has two different mechanisms: an encoder that reads the input text, and a decoder that generates a task's prediction. As BERT's goal is to build a language model, only the encoder procedure is used. The Transformer encoder reads the complete sequence of words at once, in contrast to directional models that read the text input sequentially (left-to-right or right-to-left). As a result, it is classified as bidirectional – however, it's more correct to describe it as non-directional. This attribute allows the model to effectively infer a word's context from its surroundings.

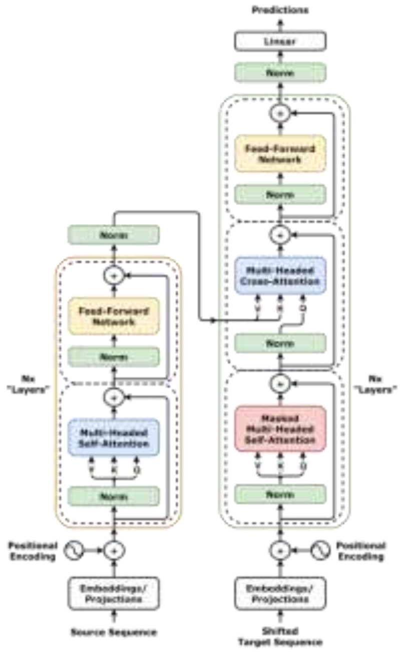

Fig. 2: Model architecture for transformers (The left and right half of this fig show how the transformer's encoder and decoder work utilizing positional embedding, Multi-Head Self/Cross Attention, and FFN respectively.)

3.3 Long Short-Term Memory (LSTM)

The LSTM, [29], [30], [31], [32], [33], [34], [35] [36], [37], [38], [39], [40], [41], [42], [43], [44] [45], [46] is an artificial Recurrent Neural Network (RNN) architecture used in deep learning? Short-term memory is a feature of these networks, and the premise here is that this feature can boost results when compared to other traditional Machine Learning approaches, [47]. Unlike standard feed-forward neural networks, LSTM has feedback connections. A typical LSTM unit consists of a cell, an input gate, an output gate, and a forget gate. The cell's gate controls the flow of information, and the cell remembers values at random time intervals as shown in Figure- 3. As an LSTM is more suited for time series analysis than other neural networks like Recurrent Neural Networks, it is chosen (RNN). Each cell in an LSTM has three types of gates that control its state: Forget Gate yields a number between 0 and 1, with 1 indicating" completely keep", while 0 designates a" completely ignore."

Memory Gate specifies which new data in the cell must be preserved. First, a sigmoid layer called the" input door layer" selects which values will be altered. The state is then updated with a vector of fresh candidate values generated by a real layer. The Output Gate determines the amount of energy generated by each cell. The cell state, as well as filtered and newly added data will decide the final value.

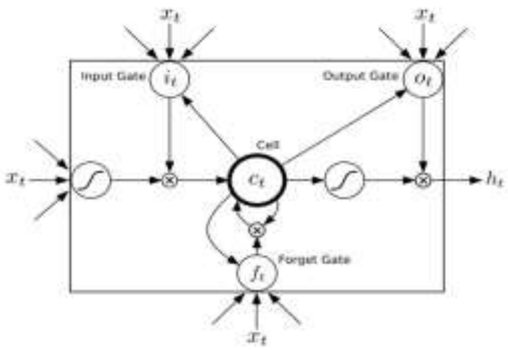

Fig. 3: Long Short-term Memory Cell.
Source: W. Commons, "Long short-term memory," *https://commons.wikimedia.org/wiki/File:Long_Short_Term_Memory.png*, 2015

3.4 The Constructed Dataset

We gathered the dataset, [4], [5], [6], [7]. from the Commercial International Bank of Egypt (COMI), and it covers stocks from February 2nd, 2012 to February 11th, 2021. The dataset contains the date of the reading, the opening price, the high and low value of the stock, news about the stock, and the volume. We had to determine the closing price for each record in these datasets because they didn't have one. We illustrate the data distribution below. We train on 70% of the dataset, while 15% is for validation, and 15% for testing. As shown below, Figure 4 illustrates the separation of the train, validation, and test sets.

Fig. 4: Data Separation

4 Results

We train on 70% of each dataset, while 15% is for validation, and 15% for testing for prediction, the input is a sample containing the last 50 days of closing prices, and the output in the prediction of the price on the next 10 days. We trained for 100 epochs with a batch size of 64 and we chose Adam - adaptive moment estimations, our optimizer. Adam is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based on training data.[48]

Table 1 shows the performance of ML algorithms on the four datasets, for the COMI dataset, the best-performing machine learning model is the Gradient Boosting Regressor with an RMSE of 0.7442 while the best-performing for the IRON data set is the Extra Trees Regressor with an RMSE 0.0451, while the best performing for ORHD data set is Gradient Boosting Regressor with an RMSE 0.1134, while the best performing for PHDC data set is Gradient Boosting Regressor with an RMSE 0.479.

Table 1. ML Algorithms Results

Algorithm	COMI	IRON	ORHD	PHDC
KNN	19.7788	2.7668	3.1308	0.9348
Decision Tree	0.9428	0.047	0.1211	0.0623
XGBOOST	0.8338	0.0609	0.1396	0.0558
Random Forest Regressor	0.7743	0.0501	0.1146	0.0489
Light Gradient Boosting Machine	0.8472	0.1071	0.2553	0.0645
Gradient Boosting Regressor	0.7442	0.047	0.1134	0.0479
AdaBoost Regressor	1.0758	0.1547	0.3064	0.0709
Extra Trees Regressor	0.7913	0.0451	0.1779	0.0491

Below table 2 shows the performance of DL algorithms on the four datasets, for the COMI dataset, the best-performing Deep learning model is the LSTM (256hu/50lag/4L) with an RMSE of

0.0259while the best-performing IRON data set is LSTM (256hu/50lag/4L) with an RMSE 0.0438, while the best performing for ORHD data set is Bi-LSTM (256hu/50lag/4L) with an RMSE 0.0054, while the best performing for PHDC data set is Bi-LSTM (256hu/50lag/4L) with an RMSE 0.01558.

Table 2. DL Algorithms Results

Architectures	COMI	IRON	ORHD	PHDC
LSTM (50hu/50lag/4L)	0.0355	0.0904	0.0083	0.0276
LSTM (256hu/50lag/4L)	0.0259	0.0438	0.0067	0.0262
Bi-LSTM (50hu/50lag/4L)	0.0433	0.0826	0.0826	0.0178
Bi-LSTM (256hu/50lag/4L)	0.0315	0.0706	0.0054	0.0156
Bi-LSTM (50hu/50lag/5L)	0.0648	0.1332	0.0091	0.0182
GRU (50hu/50lag/4L)	0.026	0.0631	0.0094	0.0163
Bi-GRU 50hu/50lag/4L)	0.029	0.0792	0.0058	0.0229
CNN-LSTM	0.0457	0.4937	0.0062	0.114
CNN-BI-LSTM	0.0478	0.404	0.0076	0.094
Attention-LSTM	0.0659	0.0615	0.0172	0.0271
Attention-BI-LSTM	0.1049	0.0625	0.0185	0.038
Attention-GRU	0.0906	0.0571	0.0201	0.0236
Attention-BI-GRU	0.0651	0.1269	0.019	0.0257

In addition, the researcher compared the performance of models that include news using FINBERT and models that do not include news in their training datasets.

Table 3. Comparison Between the Standalone model ("Numerical Data Model" combined with "FINBERT") and "Numerical Data only" Model

Numerical Data Model	News Data Model	RMSE
Bi-LSTM (256hu/50lag/4L)	None	0.0315
LSTM (256hu/50lag/4L)	None	0.0259
GRU (50hu/50lag/4L)	None	0.026
Bi-GRU (50hu/50lag/4L)	None	0.029
Bi-GRU (256hu/50lag/4L)	Bi-GRU (256)	0.0272
LSTM (256hu/50lag/4L)	LSTM (256)	0.022

The researcher also compared the effect of adding the news text as a feature with a stand-alone model using FinBERT and Numerical Data Model as shown in Table 4

Table 4. Comparison between "News Data Model" as features or labeled (positive, Negative, or neutral) and Standalone model ("Numerical Data Model" combined with "FINBERT")

Numerical Model	News Data Model	RMSE
Bi-GRU	News text as a feature	0.027
LSTM	News text as a feature	0.030
Bi-GRU	Bi-GRU	0.0272
LSTM	LSTM	0.0220

5 Discussion

Word embedding, [49], [50], [51]. refers to a group of language modeling and feature learning approaches used in natural language processing (NLP) to map words or phrases from a lexicon to real-number vectors. It can recognize a word's context in a document, its semantic and syntactic similarities, and its relationship to other words, among other things. Word embeddings are primarily utilized as input features in other models created for specific objectives. BERT has an advantage over models like Word2Vec [49] because while each word has a fixed representation under Word2Vec regardless of the context within which the word appears, BERT [26] produces word representations that are dynamically informed by the words around them. Aside from capturing obvious differences like polysemy, context-informed word embeddings capture other forms of information that result in more accurate feature representations, which in turn results in better model performance. BERT expects input data in a specific format, with special tokens A special token, [CLS], is at the beginning of our text. This token is used for classification tasks, but BERT expects it no matter what your application is. [SEP] is a special token, to mark the end of a sentence, or the separation between two sentences, in addition, we must tokenize our text into tokens that match BERT's vocabulary. BERT requires input, a series of numbers that link each input token to its index number in the BERT tokenizer vocabulary, for each tokenized sentence. As previously mentioned, the BERT base model employs 12 layers of transformer encoders, with each token's output serving as a word embedding. The BERT authors tested word-embedding strategies by feeding various vector combinations as input features to a Bi-LSTM employed on a named entity identification task and observing the resulting F1 ratings. The authors discovered that summing the final four levels was one of the top-performing options, [52].

5.1 Experiments and Results Analysis

While testing other models to train along with our proposed LSTM variant, we tried a BI-GRU. Gated recurrent units (GRUs), [53]. are a gating mechanism in recurrent neural networks. The GRU is comparable to an LSTM with a forget gate, but it does not have an output gate, hence it has fewer parameters. GRUs have been proven to perform better on some smaller and less frequent datasets. The purpose of GRU is to solve the problem of disappearing gradients in recurrent neural networks. The GRUs abandoned the cell state in favor of data transfer via the concealed state. It also has only two gates, one for resetting and the other for updating. The update gate works in the same way that the LSTM forget and input gates do.

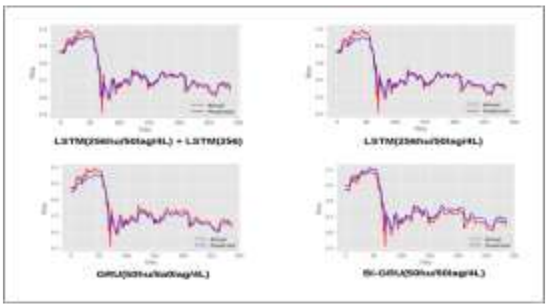

Fig. 5: Comparison between Different Architectures

A Bidirectional GRU, also known as a Bi-GRU, is a sequence processing model that consists of two GRUs, one of which takes input in one direction and the other in the other. Only the input and forgets gates are used in this bidirectional recurrent neural network. Bi-GRU works similarly to a BI-LSTM, providing more context to the network and allowing it to understand the problem faster and more completely. The Bi-GRU model yielded an RMSE of 0.0272. We also compared the effect of adding the news text as a feature and as a stand-alone model. From our observation, using a dedicated model for news sentiment analysis proved to be more effective than adding the news in one model as a labeled value. This is due to the model complexity of the stand-alone model for the news, being able to catch specific features in the text data is better than training with an additional feature that might not have much useful information compared to the numerical data.

Table 4. Comparison Between the Standalone model ("Numerical Data Model" combined with "FINBERT") Model and the "News Text as a Feature" Model

Numerical Data Model	News Data Model	RMSE
Bi-GRU	News text as a feature	0.027
LSTM	News text as a feature	0.030
Bi-GRU	Bi-GRU	0.0272
LSTM	LSTM	0.0220

In addition, we compared the performance of models that include news and models that do not include news in their training datasets. From our observations, the models that incorporated news as either a stand-alone model with FINBERT or an added feature (labeled) performed better than models that did not incorporate news. To sum up Table 1 and Table 2, the best-performing model without news is the LSTM with an RMSE of 0.0259 while the best-performing model with news is the LSTM with a stand-alone LSTM model for news yields an RMSE of 0.0220, followed by the Bi-GRU model with news as a feature with an RMSE of 0.027.

6 Conclusion and Future Work

The research proves that Deep learning models are better at catching and learning specific features that can give the edge for their prediction. In addition, the models that incorporated news as either a stand-alone model or an added feature performed better than models that did not incorporate news. Also using a dedicated model for news sentiment analysis proved to be more effective than adding the news in one model as a labeled value. LSTM-based models proved more accurate than the rest of the models, yielding the least RMSE across all datasets, followed closely by GRU-based models. Also, the results indicate that our model (FinBERT + LSTM) utilized the essential features in the news to accurately predict the state of the stock with a low error rate.

The researcher also demonstrated the effectiveness of Bert embeddings with FinBERT and LSTMs for stock market prediction with news representing a sentiment of stocks. Results indicate that our model utilized the essential features in the news to accurately predict the state of the stock with a low error rate.

In our future work, the researcher aims to test on larger datasets while testing on state-of-the-art models. also aim to perform sentiment analysis on stock market-related news which is believed could add to the robustness of the model's prediction. Also, the researcher aims to perform fake news analysis on stock market-related news which beliefs could add to the robustness of the model's prediction. Adding more financial features relevant to the customer could add more accuracy to the model prediction; select the most fitting facilities for the customer and obtain more profit for the customer. Also expanding the scope to include another investment industry rather than the stock market, such as the gold industry, petroleum industry, real estate industry …etc.

References:

[1] E. F. Fama, ‘‘Efficient capital markets: A review of theory and empirical work,’’ J. Finance, 1970, vol. 25, (2) pp: 383–417.

[2] PD. Yoo, MH. Kim, and T. Jan, “Machine learning techniques and use of event information for stock market prediction: A survey and evaluation”, In International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce Nov 28 (Vol. 2, pp. 835-841). IEEE, 2005

[3] E. Chong, C. Han, Park F.C, “Deep learning networks for stock market analysis and prediction”, Methodology: data representations, and case studies. Expert Systems with Applications. Oct 15 2017, 83:187-205, Retrieved from: https://dro.dur.ac.uk/21533/

[4] Mubashir. (n.d. a) Egyptian Iron and Steel (IRON). Retrieved from: https://english.mubasher.info/markets/EGX/stocks/IRON

[5] Mubashir. (n.d.-b). Commercial International Bank - Egypt (COMI). Retrieved from: https://english.mubasher.info/markets/EGX/stocks/COMI

[6] Mubashir. (n.d.-c). Orascom Development Egypt (ORHD). Retrieved from: https://english.mubasher.info/markets/EGX/stocks/ORHD

[7] Mubashir. (n.d.-d). Palm Hills Development Co SAE (PHDC)). Retrieved from: https://english.mubasher.info/markets/EGX/stocks/PHDC

[8] Magoč T, Modave F, Ceberio M, Kreinovich V. Computational methods for investment portfolio: the use of fuzzy measures and constraint programming for risk management. In Foundations of Computational Intelligence Volume 2 2009 (pp. 133-173). Springer, Berlin, Heidelberg.

[9] Abdullah SS, Rahman MS, Rahman MS. Analysis of the stock market using text mining and natural language processing. In2013 International Conference on Informatics, Electronics and Vision (ICIEV) 2013 May 17 (pp. 1-6). IEEE.

[10] Wang Z, Ho SB, Lin Z. Stock market prediction analysis by incorporating social and news opinion and sentiment. In2018 IEEE International Conference on Data Mining Workshops (ICDMW) 201812

[11] Kumar G, Jain S, Singh UP. Stock market forecasting using computational intelligence: A survey. Archives of Computational Methods in Engineering. 2021 May;28(3):1069-101.

[12] Shah D, Isah H, Zulkernine F. Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies. 2019 Jun;7(2):26.

[13] Chambers D, Dimson E, Foo J. Keynes the stock market investor: a quantitative analysis.

Journal of Financial and Quantitative Analysis.2015 Aug;50(4):843-68.
[14] Jorion P. The pricing of exchange rate risk in the stock market. Journal of financial and quantitative analysis. 1991 Sep;26(3):363-76.
[15] Eun CS, Shim S. International transmission of stock market movements. Journal of financial and quantitative Analysis. 1989 Jun;24(2):241-56.
[16] Chowdhury M, Howe JS, Lin JC. The relation between aggregate insider transactions and stock market returns. Journal of Financial and Quantitative Analysis. 1993 Sep;28(3):431-7.
[17] Griffith J, Najand M, Shen J. Emotions in the stock market. Journal of Behavioural Finance. 2020 Jan 2;21(1):42-56.
[18] Fataliyev K, Chivukula A, Prasad M, Liu W. Stock Market Analysis with Text Data: A Review. arXiv preprint arXiv:2106.12985. 2021 Jun 23.
[19] Nti IK, Adekoya AF, Weyori BA. A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. Journal of Big Data. 2021 Dec;8(1):1-28.
[20] Althelaya KA, El-Alfy ES, Mohammed S. Evaluation of bidirectional LSTM for short- and long-term stock market prediction. In2018 9th international conference on information and communication systems (ICICS) 2018 Apr 3 (pp. 151-156). IEEE.
[21] Fama EF. Efficient capital markets: A review of theory and empirical work. The journal of Finance. 1970 May 1;25(2):383-417.
[22] Liu G, Wang X. A numerical-based attention method for stock market prediction with dual information. IEEE Access. 2018 Dec 12; 7:7357-67.
[23] Li X, Wu P, Wang W. Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. Information Processing & Management. 2020 Sep 1;57(5):102212.
[24] Chen Y, Lin W, Wang JZ. A dual-attention-based stock price trend prediction model with dual features. IEEE Access. 2019 Oct 8; 7:148047-58.
[25] Sharma A, Bhuriya D, Singh U. Survey of stock market prediction using machine learning approach. In2017 International conference of electronics, communication and aerospace technology (ICECA) 2017 Apr 20 (Vol. 2, pp. 506-509). IEEE.
[26] Najafabadi SR. *Prediction of stock market indices using machine learning* (Doctoral dissertation, McGill University).
[27] Yoo PD, Kim MH, Jan T. Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06) 2005 Nov 28 (Vol. 2, pp.835-841). IEEE.
[28] Chong E, Han C, Park FC. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications. 2017 Oct 15; 83:187-205.
[29] Hochester S, Schmidhuber J. Long short-term memory. Neural computation. 1997 Nov 15;9(8):1735-80.
[30] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. 2014 Sep 1.
[31] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017;30.
[32] Kalyani J, Bharathi P, Jyothi P. Stock trend prediction using news sentiment analysis. arXiv preprint arXiv:1607.01958. 2016 Jul 7.
[33] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018 Oct 11.
[34] Sridhar S, Sanagavarapu S. Analysis of the effect of news sentiment on stock market prices through event embedding. In2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS) 2021 Sep 2 (pp. 147-150). IEEE.
[35] Vanstone BJ, Gepp A, Harris G. Do news and sentiment play a role in stock price prediction? Applied Intelligence. 2019 Nov;49(11):3815-20.
[36] Mate GS, Siddhant A, Rutuja K, Maitreyi M. Stock prediction through news sentiment analysis. Journal of Architecture & Technology. 2019 Aug;11(8):36-40.

[37] Names L, Kiss A. Prediction of stock values changes using sentiment analysis of stock news headlines. Journal of Information and Telecommunication. 2021 Jul 3;5(3):375-94.

[38] Liu Z. Ship adaptive course keeping control with nonlinear disturbance observer. IEEE access. 2017 Aug 21; 5:17567-75.

[39] Zhang X, Qu S, Huang J, Fang B, Yu P. Stock market prediction via multi-source multiple instance learning. IEEE Access. 2018 Sep 13; 6:50720-8.

[40] Liu H. Leveraging financial news for stock trend prediction with attention-based recurrent neural network. arXiv preprint arXiv:1811.06173. 2018 Nov 15.

[41] Wen M, Li P, Zhang L, Chen Y. Stock market trend prediction using high-order information of time series. Ieee Access. 2019 Feb 26; 7:28299-308.

[42] Ma Y, Zong L, Yang Y, Su J. News2vec: News network embedding with subnode information. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) 2019 Nov (pp. 4843-4852).

[43] Araci D. Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063. 2019 Aug 27.

[44] Zhang Y, Xu Z. BERT for question answering on SQuAD 2.0. Stanford University Report. 2019.

[45] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. Open AI blog. 2019 Feb 24;1(8):9.

[46] Moghar A, Hamiche M. Stock market prediction using LSTM recurrent neural network. Procedia Computer Science. 2020 Jan 1; 170:1168-73.

[47] Mehtab S, Sen J, Dutta A. Stock price prediction using machine learning and LSTM-based deep learning models. In Symposium on Machine Learning and Metaheuristics Algorithms, and Applications 2020 Oct 14 (pp. 88-106). Springer, Singapore.

[48] R. Zaheer and H. Shaziya, "A Study of the Optimization Algorithms in Deep Learning," *2019 Third International Conference on Inventive Systems and Control (ICISC)*, 2019, pp. 536-539, Doi: 10.1109/ICISC44355.2019.9036442.

[49] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. 2013 Jan 16.

[50] Pennington J, Socher R, Manning CD. Glove: Global vectors for word representation. InProceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) 2014 Oct (pp. 1532-1543).

[51] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with sub word information. Transactions of the association for computational linguistics. 2017 Dec 1; 5:135-46.

[52] Ronran C, Lee S, Jang HJ. “Delayed combination of feature embedding in bidirectional LSTM CRF for NER”. Applied Sciences. 2020 Jan;10(21):7557.

[53] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. 2014 Jun 3.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Mohsen Hassan formulated the research goals, Aims, and ideas for this paper, in addition to gathering and analyzing data. Hassan Also carried out the research methodology, The original draft was written by Hassan.

Amr Ghonim verified and validated the research experiment output; he also reviewed and edited the paper.

Mohsen Hassan and Amr Ghonim curated and maintained data in addition to producing metadata, mohsen Hassan carried out research investigation and data collection while Amr Ghonim managed the investigation process.

Osama Imam Managed and coordinated the research activity planning and execution

Aliaa Youssef supervised and lead responsibility for the research activity planning, execution, reviewed and validated the paper