Forecasting Stock Prices Using News Headlines

Felberg
Data Science Program (Columbian College of Arts & Sciences)
George Washington University
Washington, D.C., USA
daniel.felberg@gwu.edu

***Abstract*—The goal of this paper is to develop a time-series forecasting model that can incorporate sentiment from news headlines (i.e. *earned media*). Historically, there has been research to support that a firm's *corporate reputation* could have an impact on its financial value, but that research has largely ignored the importance of *earned media* by itself. By modelling firms' stock prices, the hope is to better establish the impact that news media can have on a company's finances. Unfortunately, due to a lack of complexity in the developed models, we are unable to establish that relationship at this time.**

Keywords*—*time series, forecasting, stock market, reputation, earned media, news headlines, sentiment analysis, LSTM

I. Introduction

The impact of news media (i.e. *earned media*) on the reputation of a firm, or company, is well-documented. To that end, it has also been observed that the reputation of a firm has tangible financial impacts on its market value, particularly in the stock market. Researchers have sought to leverage this sentiment by incorporating NLP of social media sentiment, but have largely ignored the impact that news media specifically can have on forecasting stock values of a company. This paper aims to improve upon existing time-series forecasting models to forecast firms' stock prices, and determine the viability to do using news headlines. Ultimately, we hope to answer the question of *whether or not it is possible to accurately model a firm's stock performance based on how they appear in news media*.

II. Literature Review

In order to establish the relevance of the research that was conducted, it is necessary to contextualize it. For that, the literature review should be divided into three parts. First, some terms that are specific to the field of marketing research must be defined. Second, we must prove that there is in fact a verifiable relationship between how a company appears in the media (i.e. its *reputation*) and how this impacts said company financially. From there, we can assume that an attempt to model stock prices using sentiment analysis of news headlines has validity to it. Third, to develop our time series forecast, we should then look at other models that have also incorporated sentiment analysis, and how these differ from our own goals.

Academics engaged in marketing research have defined *reputation* (more specifically, *corporate reputation*) as "Observers' collective judgments of a corporation based on assessments of the financial, social, and environmental impacts attributed to the corporation over time" [1]. One of the ways you can measure *corporate reputation* is through *earned media*. The Harvard Business School [2] defines *earned media* as "public exposure through […] **media coverage** resulting from your […] services' quality and relevancy. […] As such, *earned media* can help amplify your brand's messaging and credibility at no cost." These definitions can now help us establish the relationship between the variables to be analyzed in this paper.

While research on the effects of *corporate reputation* on a company's financial value has been substantial, many of these largely focused on overall *corporate reputation*, and less so on *earned media* alone. For instance, an examination by Alloza, Carreras, and Carreras [3] compiled literature that measured *corporate reputation* using a variety of factors, including stakeholder opinions (e.g. investors, financial analysts, etc.), as well as the mass media (which falls under *earned media*). They find "that there is a relationship between reputation and business value in its broadest sense and that it includes financial performance, market value, revenue and profit." Notably, the authors also bring attention to an older study by Deephouse [4], who developed a rating system for commercial banks using the "overall evaluation of a firm presented in the media," (i.e. *earned media*). He concluded that "media reputation is a resource that increases the performance of commercial banks." Other, more recent studies have also attempted to demonstrate how *corporate reputation* is correlated with stock prices, but, just as the previous study, used rankings as a measure of reputation. Fernández-Gámez, Gil-Corral, Galán-Valdivieso [5] used these rankings to perform a Generalized Regression Neural Network (GRNN), and found that the "Presence of firms in […] rankings has a positive influence on financial performance [and] A higher [ranking] is positively related with shares' market value." Febra, Costa, and Pereira's [6] results also corroborated those of their predecessors: "firms listed in [the] Reputation Quotient [ranking] have a […] higher reputation level [and] lower risk." In summary, existing literature on either *corporate reputation* or *earned media* affecting firms' financial outcomes overwhelmingly supports the argument that they do appear to be positively correlated. However, it should be noted that since many of the aforementioned studies utilized rankings as a measure of reputation, it's likely that some subjectivity was introduced in the analysis.

The last component of the literature review will focus on past studies that have incorporated some form of sentiment analysis into creating financial time series forecasting models. A preliminary model developed by Hassan [7] successfully forecasted the directional movement of the S&P 500 with greater accuracy compared to a benchmark model by incorporating Twitter data. The researcher used a variety of sentiment analysis tools to classify Tweets from prominent figures of the financial world, including BERT, FinBERT, and SP500BERT (a custom model trained on manually-labeled Twitter data). More recent studies have also shown promise by making use of Long Short-Term

Memory (LSTM), a deep recurrent neural network (RNN) method [8]. Using post data from a Chinese stock message board (*Eastmoney.com*), Bu, Li, Li, and Wu [8] performed sentiment analysis on those through different text classification algorithms, and then forecasted the subsequent trading day's market open prices of the CSI 300 index using various LSTM models. Gülmez [9] was also able to showcase the strength of LSTM models in forecasting stock prices by optimizing their hyperparameters with Metaheuristic algorithms, such as Artificial Rabbits Optimization algorithm (ARO). Lastly, a recent study by Asgarov [10] sought to incorporate Twitter sentiment into an LSTM model to predict closing stock values for Apple and Tesla. They found that the LSTM model that made use of those sentiment scores "can capture the complex dynamics underlying stock price movements." To conclude, recent studies have shown great promise in using LSTM models to forecast stock values, and by including sentiment analysis from external data sources.

What our literature review has not shown, however, is the degree to which we could use *earned media* to forecast stock values for a variety of companies. It is within this context that we will perform our analysis.

III. Data Collection and Pre-Processing

Having established the potential value in creating LSTM models that make use of *earned media*, we now turn to gathering and processing the necessary data. This will be done in three phases: 1) scraping news headlines that mention the specified company, 2) performing sentiment analysis on the headlines, and 3) fetching daily closing stock values for the same specified company. Once the data is collected and properly formatted, we can then start to train and test the LSTM models.

For obtaining the news headlines, a custom scraper was built that uses the Python library `Beautiful Soup` to fetch news headlines from the *Financial Times* website. Once the chosen company and its respective ticker symbol is defined, the scraper asks the user to input the number of search result pages that it will need to iterate through. Generally, it is recommended to enter a value that does not exceed `11`, or else the scraper will time out. After the scraper is executed, it returns the news headlines as a data frame, along with their respective date of publication as an index. The scraper also returns two date values: `start_date` and `end_date`, which will be used later on when fetching the company stock data.

Next, the headlines will undergo sentiment analysis. For this step, a fine-tuned version of `distilroberta-base` was used, called `distilroberta-finetuned-financial-news-sentiment-analysis`. This model was trained specifically on a dataset consisting of 4,840 financial news sentences from the English language, and therefore should provide more accurate sentiment compared to other generic text classification models [11]. Since the focus of this paper is on time series forecasting, no other sentiment analysis models were used for comparison, though new ones may be incorporated going forward to see if they impact the results. It should also be noted that the model used is publicly available on Hugging Face library [11]. After performing the sentiment analysis on the scraped headlines, two columns are added to the `headlines` data frame: `Label` and `Score`, which represent the sentiment with the highest confidence score (`positive`, `neutral`, `negative`), and the respective confidence score. The `Label` column is then re-mapped to another column, such that `positive` becomes `1`, `neutral` becomes `0`, and `negative` becomes `-1` (see Fig. 1). Lastly, the count for the most frequently occurring date index in the `headlines` data frame is obtained, as we will need it to specify the padding for the LSTM models later on.

TABLE I.

Index	Headline	Label	Score	Numeric Label
'2024-12-09'	China launches antitrust probe into Nvidia	Neutral	0.998862	0
'2024-11-21'	Nvidia's revenue nearly doubles as AI chip dem...	Positive	0.999671	1
'2024-08-29'	Nvidia brings out the crowd but not the fireworks	Neutral	0.999871	0
'2024-08-29'	Nvidia faces looming test on use of chips	Neutral	0.998440	0
'2024-12-03'	The geopolitics of chips: Nvidia and the AI boom	Neutral	0.999856	0

Fig. 1. First five rows of the `headlines` data frame (after sentiment analysis and label re-mapping).

The final step before we can begin to train and test the LSTM models on the data is to pull the daily closing stock values. For this we use the Yahoo Finance API, which returns the `stock_values` indexed list. This list has the daily closing stock values for the chosen company, with its respective date as an index. If an error is returned, it is likely due to a ticker symbol being misspelled, so make sure to check the Yahoo Finance website for the correct ticker.

At this stage we are now left with two data structures: `headlines`, which contains the scraped headlines, the sentiment numeric labels, and their respective date index, and `stock_values`, which has the daily closing stock values from Yahoo Finance.

IV. Methodology

In this paper, four LSTM models were developed, and subsequently trained and tested on the data. For all models, the sequence length, or time steps, were set to 1, the epochs were set to 20, and the batch size was set to 32. Additionally, the train-test split was set to 80-20. The architecture of the models can be seen in the tables below (Figs. 2-5):

TABLE II.

Layer	Parameter
Input	(1,1)
LSTM	25
LSTM	25
Dense	25
Dense	1

Fig. 2. First LSTM model (LSTM1) architecture (no headlines).

TABLE III.

Layer	Parameter
Input	(1,1)
LSTM	25
LSTM	25
Dense	1

Fig. 3. Second LSTM model (LSTM2) architecture (no headlines).

TABLE IV.

Layer	Parameter
Input	(1,2)
LSTM	25
LSTM	25
Dense	25
Dense	1

Fig. 4. Third LSTM model (LSTM3) architecture (with headlines).

TABLE V.

Layer	Parameter
Input	(1,2)
LSTM	25
LSTM	25
Dense	1

Fig. 5. Fourth LSTM model (LSTM4) architecture (with headlines).

For the models that incorporated the headline news sentiment (Figs. 4-5), an outer join was performed to combine the `headlines` and `stock_values` data frames. However, for dates where more than one news headline was published, those had to be concatenated in a list. Using the most common date variable from earlier, we then add a padding of `0` for all dates that require it, so that all lists have the same length. We then average the lists so the dimensions can be pipelined adequately into the LSTM models.

Each model was tested using data from the top ten global companies by market capitalization (Nvidia, Apple, Microsoft, Amazon, Alphabet (Google), Saudi Aramco, Meta Platforms, Berkshire Hathaway, TSMC, and Tesla). To avoid potential complications with news scraping, the companies "Alphabet (Google)", "Saudi Aramco", and "Meta Platforms" were respectively inputted as "Google", "Aramco", and "Facebook".

V. Results and Discussion

Due to continued variance in model performance, to account for randomness, three iterations of each model were conducted, so a more complete picture could be composed. Using RMSE to measure how well the models performed, the results are shown in Figs. 6-8, with those that performed the best for each company being highlighted:

TABLE VI.

Company	Ticker	LSTM1	LSTM2	LSTM3	LSTM4
NVIDIA	NVDA	5.0781	12.6722	5.1487	11.6729
Apple	AAPL	5.7745	12.422	4.8278	9.7018
Microsoft	MSFT	5.5537	10.6744	6.6481	8.2138
Amazon	AMZN	3.9383	7.6315	4.3874	6.5533
Google	GOOG	6.0616	11.1451	4.5245	9.2089
Aramco	2222.SR	0.2355	0.2756	0.2431	0.2641
Facebook	META	16.447	47.2765	13.8866	27.283
Berkshire Hathaway	BRK-A	9066.7334	38368.7457	9150.6053	24210.8152
TSMC	TSM	7.6242	10.5543	7.9189	9.1132
Tesla	TSLA	10.25	11.0606	10.0546	10.6955

Fig. 6. LSTM modelling results – First iteration

Based on the first iteration of the LSTM modelling, we see that half of the models that performed the best were ones that did not incorporate news headline sentiment, while the other half were models that did. Interestingly, it also seems that the LSTM2 and LSTM4 models were the worst-performing across the board, suggesting that in removing the second-to-last dense layer, this decreased model performance overall. Further testing is needed to confirm this.

TABLE VII.

Company	Ticker	LSTM1	LSTM2	LSTM3	LSTM4
NVIDIA	NVDA	5.5762	8.5023	4.5493	4.8111
Apple	AAPL	5.3571	10.5778	4.9817	9.6739
Microsoft	MSFT	5.1183	13.6263	6.297	7.7407
Amazon	AMZN	4.0844	6.5961	3.5374	6.0456
Google	GOOG	5.5509	9.1127	6.0938	8.6193
Aramco	2222.SR	0.22	0.2811	0.2213	0.2538
Facebook	META	15.537	37.2793	10.969	32.3243
Berkshire Hathaway	BRK-A	15107.4927	38363.4021	10855.5346	33679.1282
TSMC	TSM	7.2026	14.4975	8.051	11.5392
Tesla	TSLA	10.9939	11.0201	10.8459	12.0311

Fig. 7. LSTM modelling results – Second iteration

The second iteration of the LSTM modelling now appears to show some improvement in performance for the models that incorporate news headlines, with six out of ten companies' stocks performing better under that model. However, once again, LSTM2 and LSTM4 are the worst performing models, confirming yet again that adding an additional dense layer generally improved model performance. This could suggest that more complex models may be needed if we wish to improve the results. However, one more iteration should be done to finalize our findings.

TABLE VIII.

Company	Ticker	LSTM1	LSTM2	LSTM3	LSTM4
NVIDIA	NVDA	4.8533	6.3436	4.9234	10.7417
Apple	AAPL	6.0234	10.7918	5.4215	9.3888
Microsoft	MSFT	5.1725	10.7472	5.9288	9.4753
Amazon	AMZN	3.8302	7.4784	4.1637	6.3794
Google	GOOG	5.7441	9.7828	7.3071	8.4759
Aramco	2222.SR	0.2299	0.2633	0.2904	0.2534
Facebook	META	16.7271	39.8547	11.7942	27.1295
Berkshire Hathaway	BRK-A	16465.6218	36633.0754	7663.6153	22896.0987
TSMC	TSM	7.4101	11.9626	8.9927	13.4486
Tesla	TSLA	10.3694	10.7385	10.7166	10.9173

Fig. 8. LSTM modelling results – Third iteration

Interestingly, the third iteration of the LSTM modelling now shows models that don't use news headlines to be the best performing, by far, with seven out of ten companies' stocks doing better with LSTM1 specifically. Simultaneously, and for the third time, LSTM2 and LSTM4 are the worst performing models, highlighting the importance of adding the second dense layer. We must now discuss the results in more detail.

VI. Conclusion

The results from the LSTM models truly present a wide range of interpretations. In one iteration, the models using news headlines seem to perform better, but for another, the opposite seems to be the case. Ultimately, this means that the results are inconclusive, and we are unable to fully answer the question that was set out at the beginning of this research paper.

One significant takeaway, however, is that for all three iterations, the models that consistently performed the worst were those that contained less LSTM layers (LSTM2 and LSTM4). This could mean that the secret to improving the LSTM models that incorporate news media is to develop more complex models, with more layers.

VII. Future Work

Having now reached the conclusion of this paper, it is important to now acknowledge errors that were made along the way, and any lessons that can be learned from those. One aspect that could potentially improve model performance is to use alternative text classification tools beyond just `distilroberta-finetuned-financial-news-sentiment-analysis`. Another, more tedious, solution is to manually classify headlines, but this may not be a viable solution. In addition, as pointed out above, another element that could've shown improvement in the performance of the LSTM models is if models with additional layers had been used, or maybe making use of further hyperparameter tuning. If given more time, I also would have liked to incorporate additional news scrapers, which by themselves could have potentially improved the models using news sentiment. For more advanced studies, incorporating other, less-obvious, financial metrics, such as goodwill share price, intangible assets, could also offer some potential for forecasting. Lastly, one crucial area for improvement is the inconsistency in model performance, and that this was likely due to the stock data relying on the scraped news data to establish a window of analysis. In the future, a definitive date range should be decided on beforehand to limit changes in the data itself. With all of that said, it's clear that there remains work to be done in the field.

Acknowledgment

I would like to thank my colleague Liang Gao for her invaluable support throughout the arduous process of researching this topic. I would also like to acknowledge that certain sections of my code were written with the help of ChatGPT.

References

[1] Barnett, M., Jermier, J. & Lafferty, B. Corporate Reputation: The Definitional Landscape. *Corp Reputation Rev* **9**, 26–38 (2006). https://doi.org/10.1057/palgrave.crr.1550012

[2] Harvard Business School Online. (n.d.). Earned vs. paid media: What's the difference? Retrieved December 8, 2024, from https://online.hbs.edu/blog/post/earned-vs-paid-media

[3] Alloza, Á., Carreras, E., & Carreras, A. (2013). *Corporate reputation*. LID Editorial

[4] Deephouse, D. L. (2000). Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Based Theories. Journal of Management, 26(6), 1091-1112. https://doi.org/10.1177/014920630002600602

[5] Manuel A. Fernández-Gámez, Antonio M. Gil-Corral, Federico Galán-Valdivieso, Corporate reputation and market value: Evidence with generalized regression neural networks, Expert Systems with Applications, Volume 46, 2016, Pages 69-76, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2015.10.028

[6] Lígia Febra, Magali Costa, Fábio Pereira, Reputation, return and risk: A new approach, European Research on Management and Business Economics, Volume 29, Issue 1, 2023, 100207, ISSN 2444-8834, https://doi.org/10.1016/j.iedeen.2022.100207

[7] George Washington University. 2022. S&P 500 Index Price Prediction Using Twitter. Retrieved December 8, 2024, from https://scholarspace.library.gwu.edu/etd/d504rm12m

[8] Yelin Li, Hui Bu, Jiahong Li, Junjie Wu, The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning, International Journal of Forecasting, Volume 36, Issue 4, 2020, Pages 1541-1562, ISSN 0169-2070, https://doi.org/10.1016/j.ijforecast.2020.05.001

[9] Burak Gülmez, Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm, Expert Systems with Applications, Volume 227, 2023, 120346, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2023.120346

[10] Asgarov, A. (2023). Predicting Financial Market Trends using Time Series Analysis and Natural Language Processing. *arXiv preprint arXiv:2309.0013*

[11] Hugging Face. (n.d.). DistilRoBERTa fine-tuned financial news sentiment analysis. Retrieved from https://huggingface.co/mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis?text=Nvidia+faces+looming+test+on+use+of+chip