

An analysis of BERT-based model for Berkshire stock performance prediction using Warren Buffet's letters

Geyang Yu

College of Software, Nankai University, Tianjin, 300450, China

2112391@mail.nankai.edu.cn

Abstract. The objective of this study is to discover and validate effective Bidirectional Encoder Representations from Transformers (BERT)-based models for stock market prediction of Berkshire Hathaway. The stock market is full of uncertainty and dynamism and its prediction has always been a critical challenge in the financial domain. Therefore, accurate predictions of market trends are important for making investment decisions and risk management. The primary approach involves sentiment analysis of reviews on market performance. This work selects Warren Buffett's annual letters to investors and the year-by-year stock market performance of the Berkshire Hathaway as the dataset. This work leverages three BERT-based models which are BERT-Gated Recurrent Units (BERT-GRU) model, BERT-Long short-term memory (BERT-LSTM) model, and BERT-Multi-Head Attention model to analyse the Buffett's annual letters and predict the Berkshire Hathaway's stock price changes. After conducting experiments, it could be concluded that all three models have a certain degree of predictive capability, with the BERT-Multi-Head Attention model demonstrating the best predictive performance.

Keywords: Stock Price Prediction, Transformer, BERT.

1. Introduction

Warren Edward Buffett, born on August 30, 1930, is one of the best-known investors in the world, who mainly invests in stocks and funds industry. He is currently the chairman and Chief Executive Officer (CEO) of Berkshire Hathaway.

As an outstanding investor, Buffett would have access to more information about market trends and be able to synthesize all the information to make better judgments. His understanding and analysis of stock trends would be much better than ordinary investors. Study Buffett's annual letter to shareholders would be an effective way to understand Warren Buffett's views on stock market trends. Every year Warren Buffet shares his views about the market and the company's (Berkshire Hathaway) stock performance for the year. And also express his views on the investment opportunity going forward, based on the economic condition [1,2].

This study uses advanced Natural Language Processing techniques to delve into these letters, seeking correlations with Berkshire Hathaway's stock performance. Natural Language Processing is a deep learning method which represents the convergence of artificial intelligence and linguistics. This interdisciplinary field is dedicated to exploring the relationship between machines and natural languages. [3]. Compared to traditional machine learning methods, the deep learning method is able to automatically learn features from the data and improve their performance [4,5]. This study uses three

Bidirectional Encoder Representations from Transformers (BERT)-based deep learning models to predict stock performance from Warren Buffet's letters and compare the extent to which these three models are better or worse at predicting changes in stock prices.

2. Method

2.1. Dataset

The dataset for this study encompasses two primary components: Warren Buffet's annual letters to investors from 1977 to 2021 and the year-by-year stock market performance of the Berkshire Hathway from 1965 to 2021 [1]. In every letter, Buffett would analyze the situation of the company's different businesses and give possible solutions to the difficulties that the company is facing. The word count for each letter is in the range of 3,000 to 16,000 words. The data of stock market performance has two columns, year, and Annual Percentage Change in Per-Share Market Value of Berkshire [1].

2.2. Models

In this study, this work leverages three BERT-based models which are BERT-GRU model, BERT-LSTM model, and BERT-multi-head attention model.

2.2.1. Preliminary Knowledge. BERT, which means Bidirectional Encoder Representations from Transformers, is a conceptually simple and empirically powerful language representation mode [6]. A pre-trained BERT mode with just one additional output layer can handle a wide range of tasks. The purpose of using BERT is mainly to get the embedding on the Warren Buffett's letters and generate the inputs of the additional layer.

Long short-term memory (LSTM), is a kind of recurrent neural network. It aims to solve the vanishing gradient problem in the traditional Recurrent Neural Networks (RNNs) [7]. A standard LSTM unit includes four fundamental components: a memory cell, an input gate, an output gate, and a forget gate [8].

Gated recurrent units (GRU) is quite similar to LSTM, but it only comprises two gates: an update gate and a reset gate [9]. It has fewer parameters than LSTM, which means GRUs are more easily to be trained [9].

Attention mechanism is now one of the most pivotal concepts in the realm of deep learning, drawing inspiration from the cognitive attention observed in humans. By introducing attention mechanism, neural networks can automatically learn and selectively focus on important information in the input, improving the performance and generalization ability of the model [10].

Self-attention is a kind of important attention mechanism. Its basic idea is that each element in a sequential data can be associated with other elements in the sequence. It captures long-range dependencies between elements by calculating their relative importance [10].

Multi-Head Attention is an improved version of Self-attention. It is equivalent to the combinations of h different self-attention layers. It allows the model to attend to different aspects of information, improving the model's generalization capabilities [10].

2.2.2. Model Architecture. This work uses the pre-trained BERT tokenizer to tokenizes text data of every letter and converting text into tokens. Then, transmitting the tokens into the pre-trained BERT model to get the embeddings and reshape every letter's embeddings to the size of 2048×768 .

The embeddings of letters will be the input of three models.

GRU models: an input layer that takes in 768-dimensional BERT embedding with a shape of 2048×768 . The output from the input layer is forwarded through an GRU layer with 256 units. Then, the output of the GRU layer is passed through three dense layers of 128 units, 64 units, and 1 unit in that order.

LSTM models: an input layer that takes in 768-dimensional BERT embedding with a shape of 2048×768 . The output from the input layer is forwarded through an LSTM layer with 256 units. Finally, the

output of the LSTM layer is passed through three dense layers of 128 units, 64 units, and 1 unit in that order.

Multi-Head Attention models: an input layer that takes in 768-dimensional BERT embedding with a shape of 2048×768 . The output from the input layer is forwarded through an Multi-Head Attention layer with 8 heads and key dimension of 128. Then, the Multi-Head Attention layer's output undergoes sequential processing through three dense layers: first with 128 units, followed by 64 units, and finally, a single unit.

3. Experiment

3.1. Data Cleaning

Firstly, it is required to align the letters with the stock performance data. Because the year of the letter starts in 1977, while the data on stock performance starts in 1965. Entries for 1965 to 1976 would be removed from the stock performance dataset to ensure consistency. were removed from the stock performance dataset to ensure congruence in timeframes.

Secondly, use English counterparts to replace the counterparts in the letters. There is a notable anomaly in the letter. The core content of these letters is written in English, but the counterparts is not English punctuation marks. It maybe affect Natural Language Processing (NLP) analyses. So, the letters are changed into "ISO-8859-1" encoding, and implement a process to replace these punctuation marks with the English counterparts.

Finally, the embedded tables within letters are discard. It could be observed that each letter contains several tables. Some tables are describing the company's performance. Some are also describing the list of shareholders of a company and their percentage of ownership. These tables, despite their informative value, pose an obvious obstacle to the NLP analyses. The first reason is that these tables are lack of emotive content. The second reason is that the tables is consist of lines, dots, and other formatting characters. These characters would severely disrupt the NLP tokenization process. Eventually, it is decided to discard all the tables within letters through a process to detect and remove them.

3.2. Model Training

First, a pre-trained models is leveraged to process text data and extract embedding vectors. BERT imposes a constraint on the sequence length, capping it at 512 tokens. Therefore, strict padding and truncation is required to ensure compliance with BERT requirements. Then, the BERT model is leveraged to convert the tokens into BERT embedding vectors and reshape the result corresponding to each letter from their original shape of $1 \times 2048 \times 768$ to a new shape of 2048×768 .

Second, the BERT embedding vectors would be the input of the three sequential models described in Section 2.2.2. All the models have been configured with the Adam optimizer, utilizing a learning rate of 10^{-5} , and the MSE (Mean Squared Error) serves as the loss function. In addition, the MAE (Mean Absolute Error) is used as a monitoring metric during training.

In the subsequent step, the models will undergo training for 500 epochs with a batch size of 16. Moreover, 10% of the training data will be partitioned for validation purposes to monitor the models' performance.

4. Results

After the model training, there are three models which have the inputs of BERT embedding and the output of predicted stock performance. In order to visualize the training results, the BERT embedding takes all the 45 letters as input. These three models for stock prediction and resulted in the following Figure 1, Figure 2, and Figure 3 respectively.



Figure 1. Result of BERT-Multi-Head Attention (Figure Credits: Original).

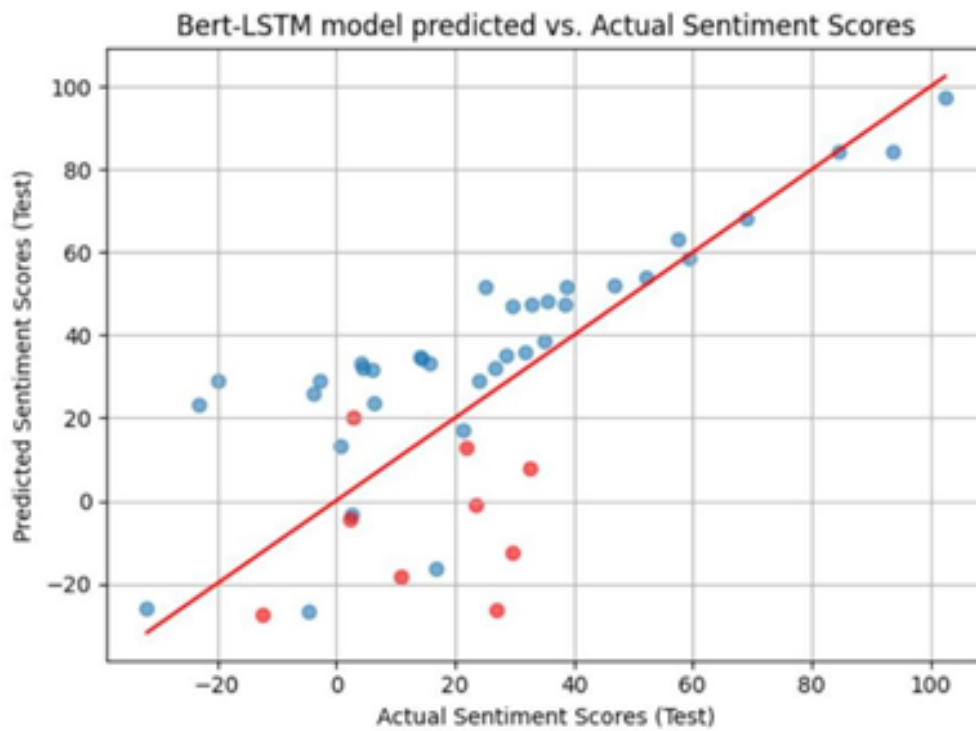


Figure 2. Result of BERT-LSTM Attention (Figure Credits: Original).

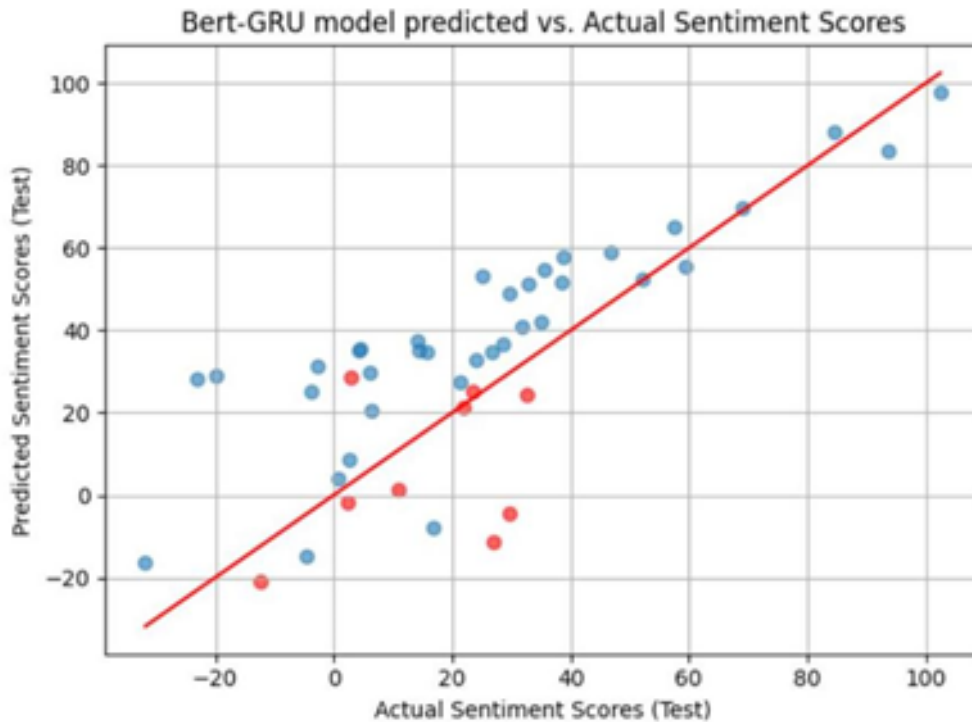


Figure 3. Result of BERT-GRU Attention (Figure Credits: Original).

In these figures, the blue dots denote predictions made on the training data, while the red dots represent the prediction in test data. The coordinates of each dot (x,y) , where x represents the actual stock performance data and y represents the predicted result. The functional expression of the red line is $y = x$, which also means that the closer a dot is to the red line, the smaller the prediction error represented by that dot.

To rigorously compare the predictive abilities of these three models, this work will use the Mean Absolute Error (MAE) of the test data as the evaluation metric. Table 1 displays the Mean Absolute Error (MAE) for the three models.

Table 1. MAE performance of various models.

Model Name	MAE
BERT-Multi-Head Attention	28.0880
BERT-LSTM	29.6538
BERT-GRU	31.5577

5. Discussion

In these experiments, some observations could be made. The BERT-multi-head attention model demonstrated the highest accuracy in predicting stock market trends among these three models. This discovery carries significant implications for the field of finance and investment decision-making.

From the graphical representation of the experimental results, it could be seen that all three models have a good fit and as can be seen from the table describing MAE error, the BERT-multi-head attention model's predictions are more accurate.

The implications of the findings are substantial. Accurate predictions of stock market trends have long been considered a critical task because it can guide investment decisions and inform risk management strategies. This research suggests that the BERT-multi-head attention model shows significant potential in stock market forecasting. Traditional stock market analysis often relies on

historical trends and various data indicators. Meanwhile, BERT models excel at extracting information from text of stock market analysis. This can be highly valuable in understanding market dynamics.

However, it's crucial to recognize the limitations of this study. The experiment is mainly based on the review of Buffet. While Warren Buffett is an exceptional investor, he is not immune to making analytical mistakes. Meanwhile, the prediction of the proposed models still has a significant margin of error. These errors can potentially lead to catastrophic consequences when guiding market investments. Additionally, the trends in the stock market are also uncertainty due to various unexpected external events. This poses challenges to predictions.

6. Conclusion

Warren Buffett's letters to investors can provide valuable insights into his views on market trends. By analyzing the sentiment of the text, this work can potentially predict changes in the stock market. The aim of this study is to find and validate effective BERT-based models for stock market prediction. Through in-depth analysis of textual data, this research has demonstrated the potential of BERT-based models in the stock market forecasting.

This paper highlights the potential of the BERT model in stock market prediction. While further optimization of these experimental methods is still needed, the finding provides new ways for applying natural language processing technology in financial analysis and decision-making. Through in-depth analysis of extensive textual data, this research has demonstrated the potential of BERT models in the realm of stock market forecasting.

The experimental result indicate that BERT-based models have the excellent capabilities in predicting the stock performance of Berkshire Hathaway based on Warren Buffet's reviews. Furthermore, among the three models experimented in this study, the BERT-multi-head attention model exhibited the best predictive performance.

Unlike traditional analysis methods based on technical indicators and historical price data, this text-based prediction method offers a new perspective on market forecasting.

However, it is crucial to emphasize that predicting the stock market remains a complex and challenging task, given the difficulty in capturing all influencing factors. Consequently, models still carry errors and limitations. Future research directions should focus on further optimizing NLP models to enhance their performance in stock market prediction. Additionally, expanding the dataset and considering more investors' reviews will contribute to a more comprehensive understanding of market dynamics.

References

- [1] Warren buffet letters to investors 197-2021, URL: <https://www.berkshirehathaway.com/letters/letters.html>. Last accessed 2023/08/22
- [2] Hu, Z., Zhao, Y., & Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1), 9.
- [3] Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5), 544-551.
- [4] Shinde, P. P., & Shah, S. (2018). A review of machine learning and deep learning applications. In 2018 Fourth international conference on computing communication control and automation, 1-6.
- [5] Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685-695.
- [6] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [7] Graves, A., & Graves, A. (2012). Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, 37-45.
- [8] Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural computation*, 12(10), 2451-2471.

- [9] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
- [10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Advances in neural information processing systems*, 30, 1-11.