Healthcare Sentiment Analysis for Market Trends

Johann Tin Tsin Fong
Simon Fraser University
jtintsin@sfu.ca

Emily Wong
Simon Fraser University
ewa47@sfu.ca

Abstract

This project explores the use of healthcare sentiment analysis for stock market predictions using natural language processing (NLP) techniques. The objective is to assess the impact of news on stock prices in the healthcare sector by extracting key entities and analyzing sentiment. The approach integrates Named Entity Recognition (NER) to identify pharmaceutical company names, drug names, healthcare organizations, and other pertinent keywords from news articles, ensuring a focus on healthcare-specific articles for sentiment classification.

To achieve this, we use an existing BERT model and fine-tune it to give sentiment scores. The sentiment classification results are compared against FinBERT, an established financial sentiment analysis model. The resulting sentiment scores are then used as inputs for a Long Short-Term Memory (LSTM) (Fischer and Krauss, 2018) model to predict stock price movements.

The effectiveness of our fine-tuned model combined with LSTM is evaluated against two baselines: (1) FinBERT + LSTM, and (2) a standalone LSTM model without sentiment input. Predictions are assessed based on their correlation with actual stock price trends in the healthcare sector.

By combining NER-driven sentiment analysis with deep learning-based stock predictions, this project aims to generate trend line predictions and offer clearer insights into how events impact stocks in the healthcare sector.

1 Introduction

This project explores healthcare sentiment analysis for stock market predictions using NLP techniques. The objective is to assess the impact of news on stock prices in the healthcare sector. Our approach begins with filtering relevant healthcare news articles from NEWSDATA.IO (NewsData.io, 2025), which are then processed by a custom NER model to extract key entities such as pharmaceutical company names, drug names, healthcare organizations, and other pertinent keywords. These articles are then manually labeled with sentiment scores, which would be used as our training dataset.

We selected this project because news articles are just one of many important factors that influence market prices. For instance, UnitedHealth's stock experienced a sharp decline following the shocking news of its CEO's assassination, resulting in billions of dollars in lost market value (Bedigan). This event underscores how unexpected occurrences can significantly impact stock movements, a phenomenon also analyzed in studies such as *Impact of Major Health Events on Pharmaceutical Stocks* (Maleki and Ghahari, 2024).

We fine-tune existing BERT-based model from HuggingFace (Wolf et al., 2020b) to give sentiment scores and compare its performance against FinBERT. FinBERT is a BERT-based NLP model trained on financial texts to classify sentiment (positive, negative, or neutral) for stock market analysis (Araci, 2019).

These sentiment scores, along with historical stock data from Alpha Vantage API (Alpha Vantage Inc., 2024), are then fed into two deep learning models, LSTM + FinBERT and LSTM + HBERT, to predict stock price movements. By combining NER-driven healthcare sentiment analysis with deep learning-based stock predictions, this project aims to generate trend line predictions and offer clearer insights into how events impact stocks in the healthcare sector.

2 Related Works

FinBERT is a BERT-based NLP model trained on financial texts to classify sentiment as positive, negative, or neutral, making it useful for stock market analysis (Araci, 2019). By analyzing financial news and reports, FinBERT helps assess market sentiment and predict stock price movements.

BioBERT, on the other hand, is designed for biomedical NLP tasks and trained on large-scale biomedical corpora like PubMed abstracts and clinical (Lee et al., 2019) notes. Unlike FinBERT, BioBERT does not perform sentiment analysis but excels at tasks like Named Entity Recognition (NER) and relation extraction, making it useful for identifying healthcare-related entities such as drug names and diseases.

Since our model will be trained on financial news, we compare its performance in healthcare sentiment analysis against FinBERT. While FinBERT is designed for financial text, it is not specifically trained on healthcare-related financial news. By fine-tuning our model on healthcare-focused financial news, we aim to improve sentiment classification and better capture the impact of medical events on stock prices.

3 Approach

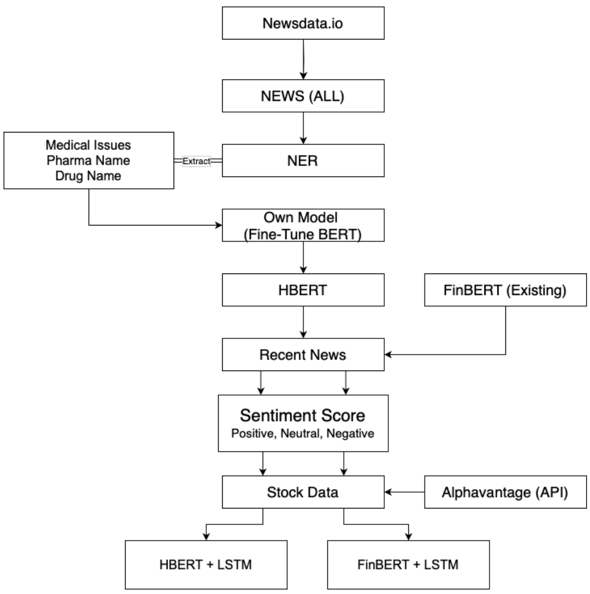

Figure 1: Architecture

For the project, we began by collecting news articles from NEWSDATA.IO and historical stock data with Alpha Vantage API. The news dataset is filtered for English-only articles, and 200 articles from each topic are randomly sampled. Then NER is applied to the processed dataset to extract healthcare-related entities, including pharmaceutical company names, medicine names, medical issues, insurance company names, and health organizations. This helps identify healthcare related news for the sentiment labeling.

Then sentiment score are manually added for each articles. These data are then used to train our own model HBERT (Healthcare BERT). Once the model has been trained, the recent news dataset is fed to both the FinBERT and HBERT model for sentiment score labeling.

Finally, we feed the sentiment score and stock data to two separate deep learning models, HBERT+LSTM and FinBERT+LSTM. These two models will then predict the stock price based on the sentiment score.

4 Experiments

4.1 Data

For the news articles, we downloaded free news datasets from NEWSDATA.IO. We gathered news on various topics, including COVID-related news, world politics, health, science and technology, and recent events. Except for the recent news, all other topics include articles from 2021.

For stock-related news and price data, we used the Alpha Vantage API, which provides real-time and historical stock market data, including financial news relevant to specific tickers.

To represent the healthcare sector in our analysis, we selected the Healthcare Select Sector SPDR Fund (ticker symbol: XLV). XLV is an ETF that tracks the performance of the Health Care Select Sector Index. It includes a diversified set of stocks from various industries within the healthcare sector, such as pharmaceuticals, biotechnology, medical devices, and health insurance (xlv). XLV was chosen as it broadly reflects movements across the healthcare industry, making it a suitable benchmark for evaluating the impact of healthcare news sentiment on stock trends.

5 NER

We implemented custom_NER by adapting the custom NER tutorial (Elsayed, 2021) into a Python script. The NER model is trained using a custom annotation file, which generates the training data required to build a custom spaCy NER model. The training results are presented in Table 1. From this table, we observe that some steps achieved perfect F1 score, precision, and recall. This may be due to overfitting, as the training and validation sets are the same.

The NER_extraction script will take the best-trained model and load it as a spaCy object for label extraction. With these entities, we will be able to better identify healthcare-related news articles.

Table 1: NER Model Training Metrics

Epoch	#	Loss Tok2Vec	Loss NER	F1 (ENTS_F)	Precision (ENTS_P)	Recall (ENTS_R)	Score
0	0	0.00	43.45	0.00	0.00	0.00	0.00
1	200	170.70	1780.56	48.55	50.18	47.02	0.49
2	400	3209.06	1686.26	55.38	69.50	46.03	0.55
4	600	1589.97	1175.38	85.38	86.17	84.60	0.85
7	800	233.17	651.67	90.48	91.25	89.74	0.90
10	1000	424.34	468.80	95.67	94.36	97.02	0.96
14	1200	863.40	349.70	96.78	96.54	97.02	0.97
19	1400	181.93	295.76	97.18	97.34	97.02	0.97
26	1600	101.28	169.77	99.50	99.67	99.34	1.00
33	1800	154.50	174.12	99.25	99.67	98.84	0.99
43	2000	791.85	130.43	99.50	99.67	99.34	1.00
55	2200	84.29	58.67	100.00	100.00	100.00	1.00
69	2400	112.01	50.31	99.92	100.00	99.83	1.00
83	2600	294.27	95.26	100.00	100.00	100.00	1.00
97	2800	87.82	44.38	99.50	99.34	99.67	1.00
112	3000	257.31	100.12	100.00	100.00	100.00	1.00
126	3200	449.72	169.88	99.67	99.67	99.67	1.00
140	3400	486.93	212.10	100.00	100.00	100.00	1.00
155	3600	31.35	10.50	100.00	100.00	100.00	1.00
169	3800	6.42	2.06	100.00	100.00	100.00	1.00

6 HBERT

We developed HBERT (Healthcare BERT) using the Hugging Face Transformers library (Wolf et al., 2020a), built on the bert-base-uncased checkpoint using Bert Model and BertTokenizer. BERT is fine-tuned for sentiment classification in healthcare-related financial news, incorporating domain-specific entity information such as company names, drug mentions, and healthcare policies into the classification process. This enables more context-aware interpretation of stock-related articles, with the model predicting article sentiment as positive, neutral, or negative.

6.1 Entity-Aware Encoding

To incorporate domain knowledge, each input sample includes a set of named entities (e.g., company names, drug names, etc.). An entity mask is constructed and aligned with tokenized input to highlight the positions of entities within the text. This binary mask is passed through an embedding layer and then averaged to form an entity-aware representation. By embedding this information alongside standard BERT embeddings, the model can better associate specific entities with their contextual sentiment.

7 LSTM

LSTM for Financial Time-Series Forecasting: This model, inspired by Fischer and Krauss, uses Long Short-Term Memory (LSTM) networks to predict stock prices based on historical data.

7.1 Dropout Layer

Dropout is a regularization method that randomly deactivates neurons during training to prevent overfitting. Applying dropout after each Bidirectional LSTM layer helps the model learn general patterns instead of memorizing data. Tuning the dropout rate balances model complexity and generalization.

7.1.1 Attention Layer

The Attention Layer improves the model by focusing on the most relevant parts of the input sequence. It assigns weights to each time step, enhancing prediction accuracy and interpretability.

7.1.2 Bidirectional LSTM

The Bidirectional (Schuster and Paliwal, 1997) LSTM processes sequences in both forward and backward directions. This dual view captures more context and improves pattern recognition, which is especially useful for modeling complex stock movements.

8 Results

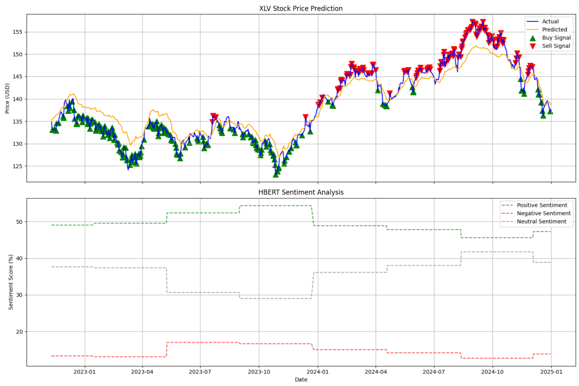

Figure 2: Stock price prediction and sentiment analysis using HBERT. The upper plot shows actual vs. predicted prices with buy/sell signals; the lower plot shows HBERT's sentiment scores over time.

Figures 2 and 3 compare the performance of HBERT and FinBERT on XLV stock price prediction, along with their corresponding sentiment analyses. In the top panels of both figures, we observe that HBERT yields predictions more closely aligned with actual price trends and produces more coherent buy/sell signals. In contrast, FinBERT shows larger deviations between predicted and actual prices, along with less consistent signal timing. The lower panels illustrate a key difference in sentiment behavior: HBERT's sentiment scores remain relatively stable over time, while FinBERT's sentiment fluctuates sharply and frequently, with abrupt transitions between positive, neutral, and negative classes.

This sentiment stability in HBERT likely stems from its domain-specific fine-tuning, which helps

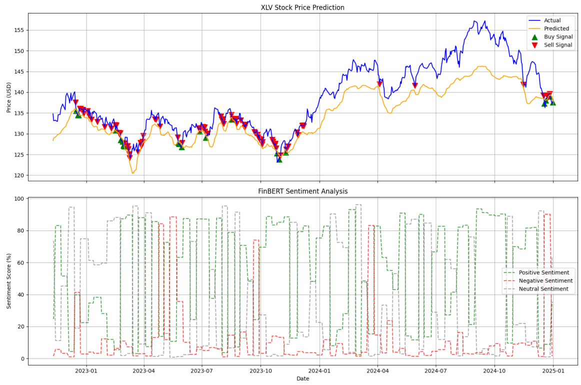

Figure 3: Stock price prediction and sentiment analysis using FinBERT. The upper plot shows actual vs. predicted prices with buy/sell signals; the lower plot shows FinBERT's sentiment scores, which are more volatile compared to HBERT.

it better contextualize healthcare-related financial news. Interestingly, HBERT's sentiment output is predominantly positive across the evaluation period. One possible explanation for this skew is the nature of its training data, which includes articles from the COVID-19 era—a time when healthcare news coverage was significantly more negative. As a result, more recent news may be interpreted as relatively positive in comparison, introducing a potential baseline shift in sentiment classification. This highlights the importance of aligning the temporal context of training data with the target prediction window to avoid bias or drift in sentiment trends. Overall, the visual comparison supports the conclusion that HBERT provides more stable and contextually relevant sentiment signals, contributing to improved stock prediction performance in the healthcare sector.

Table 2: Price Movement Prediction Metrics (2024)

Model	Accuracy	Precision	Recall
LSTM-FinBERT	0.4963	0.4981	0.4963
LSTM-HBERT	0.5336	0.5323	0.5336
LSTM-only	0.5392	0.5388	0.5392

As shown in Table 2, the LSTM-only model achieved the highest accuracy (53.92%), slightly outperforming both sentiment-enhanced models. The LSTM-HBERT model, which uses sentiment scores from our fine-tuned healthcare-specific BERT, outperformed the LSTM-FinBERT baseline with an accuracy of 53.36% compared to 49.63%. This suggests that domain-specific sentiment analysis (HBERT) provides more relevant signals for stock prediction in the healthcare sector than general financial sentiment (FinBERT), though the overall performance gains remain modest.

Table 3: F1 Score for Price Movement Prediction (2024)

Model	F1 Score
LSTM-FinBERT	0.4971
LSTM-HBERT	0.5324
LSTM-only	0.5382

8.1 F1 Score Analysis

Table 3 presents the F1 scores for each model, offering a balanced view of precision and recall. The LSTM-only model achieved the highest F1 score at 0.5382, indicating that even without sentiment input, the model captures price movement patterns relatively well. The LSTM-HBERT model closely follows with an F1 score of 0.5324, outperforming the LSTM-FinBERT model (0.4971). This further supports the value of healthcare-specific sentiment analysis over general financial sentiment, though the margin between sentiment-based and non-sentiment-based models remains relatively narrow.

Table 4: Comparison of Model Performance (2024)

Model	MAE	R-squared
LSTM-FinBERT	4.1255	0.6475
LSTM-HBERT	2.9541	0.7842
LSTM-only	1.7570	0.9238

8.2 Regression Performance

Table 4 reports the regression metrics—Mean Absolute Error (MAE) and R-squared—for each model. The LSTM-only model demonstrated the strongest performance, with the lowest MAE (1.7570) and the highest R-squared value (0.9238), indicating high predictive accuracy and strong correlation with actual stock price trends. The LSTM-HBERT model showed notable improvement over LSTM-FinBERT, achieving a lower MAE (2.9541 vs. 4.1255) and a higher R-squared (0.7842 vs. 0.6475). These results confirm that while healthcare-specific sentiment (HBERT) enhances predictive power compared to general financial sentiment (FinBERT), models without sentiment input (LSTM-only) still yield the most accurate stock price predictions in this setting.

9 Conclusion and Future Work

This project explored the use of healthcare-specific sentiment analysis for predicting stock price movements in the healthcare sector. Our fine-tuned BERT model (HBERT) outperformed FinBERT in both classification and regression tasks. However, the LSTM-only model, which used no sentiment input, achieved the best overall performance across all metrics.

These results suggest that sentiment analysis, while helpful, may not be a dominant predictor for short-term price movement in the healthcare sector. Many healthcare news articles are long-term in nature—focusing on policy, research, or clinical trials—and do not trigger immediate market reactions. Moreover, XLV, the ETF used in this study, includes large, stable companies like JNJ, PFE, and UNH, which may be less sensitive to daily headlines. News relevance and timing also vary, and sentiment may not always align with the ETF's constituents. In contrast, the LSTM-only model likely captured short-term technical and momentum-based signals that drive ETF movements more effectively.

Future work could enhance sentiment alignment by filtering for ticker-specific relevance and modeling news-to-price time lags. Integrating technical indicators and using multimodal inputs like trading volume or volatility measures may also improve prediction accuracy. Expanding analysis to individual healthcare stocks rather than the ETF level may reveal more direct sentiment-price relationships.

The custom NER model can be further improved by expanding the custom annotation to include a wider range of articles, which would help to improve its accuracy. In particular, some terms are incorrectly labeled as medicines or medical illnesses when they should not. This issue may stem from overfitting, as the training and evaluation datasets for the NER model were identical, potentially limiting the model's ability to generalize.

References

Healthcare select sector spdr fund (xlv) company profile & overview. `https://finance.yahoo.com/quote/XLV/profile/`. Accessed: 2025-04-10.

Alpha Vantage Inc. 2024. Alpha vantage api for stock market data. `https://www.alphavantage.co`. Accessed: 2025-03.

Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint*, arXiv:1908.10063.

Mike Bedigan. Unitedhealth lost $63 billion in value after ceo's death.

Amr Elsayed. 2021. Custom ner with spacy 3. Accessed: 2025-03-18.

Thomas Fischer and Christopher Krauss. 2018. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669.

Jinhyuk Lee et al. 2019. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, arXiv:1901.08746.

Morteza Maleki and SeyedAli Ghahari. 2024. Impact of major health events on pharmaceutical stocks. *arXiv preprint*, arXiv:2408.01883v1.

NewsData.io. 2025. Datasets from newsdata.io. Accessed: 2025-03-05.

Mike Schuster and Kuldip Paliwal. 1997. Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on*, 45:2673 – 2681.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. 2020a. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45. Association for Computational Linguistics.

Thomas Wolf et al. 2020b. Transformers: State-of-the-art natural language processing. *arXiv preprint*, arXiv:2005.14165.