ML Project Report

Yuhong LI, Artur Garipov,
Alessandro Ivashkevich

March 11, 2025

1 Abstract & Introduction

In this project we will integrate sentiment analysis with some algorithms for the forecasting of financial time series.

For a long time, forecasting stock market prices has become an attractive investigative topic to both analysts and researchers, due to its significant role in economic world as well as enormous potential benefits it offers. However, stock prices are hard to predict because of their highly volatile nature, influenced by a wide range of factors such as political and economic conditions, shifts in leadership, investor sentiment, and various other elements.

2 Problem Definition

2.0.1 Problem Definitions

For every stock analyzed, there are:

- A financial time series $\mathbf{y_s} = \{y_{1s}, y_{2s}, \ldots, y_{Ts}\}$, where $y_t \in R$ represents the observed financial value at time t for the stock s.
- A set of inputs $\mathbf{X} = \{\mathbf{x}_{1s}, \mathbf{x}_{2s}, \ldots, \mathbf{x}_{Ts}\}$ at time t for the stock s.
- Sentiment scores $\mathbf{Z}$= $\{\mathbf{z}_{1s}$, $\mathbf{z}_{2s}$,$\ldots$, $\mathbf{z}_{Ts}\}$ at time t for the stock s.
- A predicted value at time t: $\hat{y}_{ts}$

The analysis will be made on daily basis, so each t will refer to a different financial day.

2.0.2 Models used

To predict the close price, we opted to use XGBoost and BI-LSTM.

XGBoost (Extreme Gradient Boosting) is a scalable, high-performance implementation of gradient-boosted decision trees, designed to optimize predictive accuracy through iterative learning. It combines shallow decision trees into a robust ensemble using regularization and parallel processing for top performance on structured data. Its ability to capture non-linear relationships and prevent overfitting (via L1/L2 penalties) makes it ideal for volatile financial markets.

Bi-LSTM (Bidirectional Long Short-Term Memory) is a type of recurrent neural network (RNN) architecture that can capture temporal dependencies in sequential data. A Bidirectional LSTM processes sequences using two LSTMs: one for forward input and the other for backward input, providing enhanced context from both past and future observations. This makes Bi-LSTMs well-suited for time series tasks with temporal dependencies.

Financial time series often exhibit nonstationary patterns and shifting trends. Bi-LSTMs effectively capture long-range dependencies and cyclical behaviors, adapting to changes directly from the data without needing pre-defined lag structures or stationarity transformations.

DeBERTa (Decoding-enhanced BERT with disentangled attention) is a state-of-the-art neural network model that extends the foundational architectures of BERT and RoBERTa. Introduced by He et al. (2021), DeBERTa improves upon its predecessors by employing an enhanced attention mechanism and an advanced masked decoder, allowing it to better capture complex dependencies in textual data.

The key innovation in DeBERTa lies in its disentangled attention mechanism, which separates the content embeddings from position embeddings, allowing the model to focus more effectively on the semantic meaning of words irrespective of their positional context. Additionally, the use of a mask decoder improves training efficiency and accuracy by dynamically predicting the masked tokens based on richer contextual representations. These advancements make DeBERTa particularly effective for a wide range of natural language processing tasks, from classification to generation, by adapting its final layer (commonly referred to as the "head") to the specific requirements of the problem at hand.

2.0.3 Constraints

First, we assume that we have complete financial time series data. This ensures we can perform our analysis without gaps in the data that could affect our results.
For simplicity in our analysis, we assume there are no missing values in financial features (X), and target variables (y). Regarding the sentiment scores Z, we have scores whenever the news are published, this means that for some financial days might not exist news for our stocks in the dataset.
Finally, we make a critical assumption about temporal alignment: the sentiment features (z_t) and financial features (x_t) are perfectly synchronized in time. This means that for any given time step t, both types of features correspond to the same moment, ensuring that our analysis captures the true relationship between sentiment and financial signals without any temporal misalignment that could confound our results.

2.0.4 Objective Functions.

Both models aim to minimize a predefined loss function, in particular we observe at the RMSE between the predicted and actual values:

$$\mathcal{L}(\theta) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2} \tag{1}$$

RMSE (Root Mean Square Error) is preferred over MSE for its interpretability, as it expresses errors in the original units by taking the square root. This is particularly useful for comparing prediction accuracy across stocks with different price ranges.

To ensure comparability, we normalized RMSE by dividing it by the range of the time series, making it scale-independent and suitable for diverse datasets.

3 Related Work

3.0.1 Some of the previous work

- "Stock Price Prediction Using **LSTM Neural Network**" by Fischer and Krauss (2018)
  The research paper presents compelling evidence supporting the effectiveness of Long Short-Term Memory (LSTM) networks in the domain of daily stock price prediction. Through comprehensive analysis, the study demonstrates that LSTM models achieve superior predictive performance compared to conventional time series approaches like ARIMA (Autoregressive Integrated Moving Average).

- "**ARIMA** Model for Forecasting the Stock Price Index" by Al-Shiab and Al-Zibdeh (2020).
  This paper analyzes ARIMA models for stock price forecasting, highlighting their strength in capturing linear patterns and autocorrelations. ARIMA performs well under specific conditions, such as stable statistical properties or regular patterns, but is less versatile than machine learning models.

  Since our goal is to handle diverse market conditions, we chose not to use ARIMA for our forecasting.

- "Stock Price Forecasting Using **XGBoost Model**" by Benteş and Horobet (2018) This paper offers valuable validation of XGBoost's strengths in financial forecasting applications. The research demonstrates how XGBoost's architectural design makes it particularly well-suited for financial data challenges - its tree-based approach naturally handles missing values without requiring imputation, while its regularization techniques help maintain model stability even when faced with noisy market data and statistical outliers.

- **Sentiment analysis and deep learning** for stock price prediction" by Zhang et al. (2019)
  The study explores Bi-LSTM models where financial news sentiments are encoded as features alongside historical price data.This paper gives a strong foundation in our project's aim. Besides the implementation of sentiment scores in Bi-LSTMs, we will experiment their implementation in XGBoost.

- **On Assessing the Performance of LLMs for Target-Level Sentiment Analysis in Financial News Headlines** (2025)
  The research paper examines the effectiveness of various sentiment analysis techniques, including traditional models, fine-tuned transformer-based models, and zero-shot large language models, in analyzing financial news headlines at the target level. It compares the performance of models such as VADER, DistilFinRoBERTa, fine-tuned DeBERTa-v3, and LLMs like ChatGPT and Gemini. The findings show that fine-tuned DeBERTa-v3 achieves the highest accuracy, while LLMs like ChatGPT-4 and Gemini 1.5 also perform well without requiring task-specific fine-tuning. The study demonstrates the effectiveness of the fine-tuned DeBERTa-v3 model for classifying news headlines into positive, negative, or neutral sentiments. Utilizing a dataset of annotated financial news articles, the model outputs probabilities for each sentiment category, facilitating nuanced sentiment estimation. The results highlight DeBERTa's potential for applications in financial analysis.

4 Methodology

For the reasons described in the section *Problem Definitions*, we built some forecasting price models, with an implementation of sentimental analysis. We will check if implementing sentiment analysis as model features will optimize their predictions or not. We used:

1. Financial time series of some influential NASDAQ titles, provided by Yahoo Finance

2. A financial news (up to 07/20) dataset

Since the financial news does not cover the news after July 2020, we will not consider the time series after that date.

4.0.1 Data collection process

The time series were downloaded throught *yfinance*. Since we are consider several variables, we collect the data-frames in different lists/dictionaries, in particular:

1. 'stocklist': list of tickers names. We selected stocks which appeared more frequently in the financial news dataset

2. 'series_stock': list of time series'data frames.
3. 'train_data' and 'test_data': dictionaries where collect repectively the training and test data from each stock
4. 'X_train_dict' , 'y_train_dict', 'X_test_dict', 'y_test_dict': subframes obtained from train_data' and 'test_data' After that, we have added some technical indicators:

4.0.2 Technical Indicators and Sentiment Probabilities

To train our model and understand the stock trends, we implemented 3 technical indicators:

- The *Moving Average Convergence Divergence* [MACD] is a widely used technical indicator in trading. It helps identify momentum and trend changes in the price of an asset. The MACD is calculated by subtracting the 26-period EMA from the 12-period EMA, which produces the MACD line. A signal line, which is the 9-period EMA of the MACD line, is then plotted on top of the MACD line.
- The *Relative Strength Index* [RSI] is a momentum oscillator that measures the speed and change of price movements. It oscillates between zero and 100, typically used to identify overbought or oversold conditions in a market.
- *Bollinger Bands* consist of three lines that move with the price of an asset: a middle band (which is typically a 20-day simple moving average) and two outer bands that are placed two standard deviations above and below the middle band.

As we will demonstrate later, we have performed the models with and wihtout sentiment scores, to understand how they improve or not the models. In particular, we utilized DeBERTa-v3, a 142-million-parameter model pre-trained on financial news text and available on the Hugging Face platform. We fine-tuned it using a labeled dataset with three sentiment categories: negative, positive, and neutral. The predict_proba function was adapted to predict probabilities for these labels, ensuring they sum to one. The fine-tuned model produced interpretable sentiment probabilities for each news article, providing insights into sentiment distribution. So, for the sentiment scores we are going to add three columns: *prob_neg, prob_pos, prob_neutr* - for each day in which appears a stock news owns probabilities are assigned.

4.0.3 User Interface for Visualization

For better understanding and practical use of our sentiment model, we developed a user-friendly interface using Flask and React to display sentiment analysis of press releases and manage submissions. The front-end uses a modular React.js architecture with React Router, SCSS, Material UI, and ECharts for dynamic routing and rendering. The Flask backend processes client requests and serves JSON data, while MongoDB handles efficient data storage. The structure is as follows:

- **Backend**
  - **mongo_db.py**: Interacts with MongoDB for database operations. Functions include `save_to_mongo` (inserts press release data and sentiment information), `search_one`, and `delete_one` (fetches or deletes data by `ObjectID`).
  - **sentModel.py**: Contains the sentiment model to process press releases and output sentiment labels and dictionaries.
  - **server.py**: Implements Flask and provides four routes:
    - `/(Index Route)`: Displays a welcome message or fetches a single database record based on `_id`.
    - `/all`: Fetches all data for display on the front-end.
    - `/submit`: Accepts a user-submitted text, processes it through the sentiment model, and stores the results in MongoDB.
    - `/delete`: Deletes a specific record based on `_id`.
- **Front-end**
  - **src**
    - **components**: Contains reusable React components.
    - **pages**: Includes main pages:
      - **Home**: Accepts user input for analysis.
      - **History**: Displays submission records.
      - **Document**: Visualizes sentiment analysis results with word clouds and highlights.

The APP offers two key features:

1. **Record Management**: Provides intuitive tools to manage press release submissions efficiently.
2. **Sentiment Visualization**: Displays positive and negative sentiment words using interactive word clouds and highlights, enhancing clarity and user experience.

5 Evaluation

5.1 Evaluation of Neural Network Models for Sentiment Analysis

To evaluate our sentiment analysis approach, we experimented with two neural network models from Hugging Face: **DeBERTa-v3** and **BERT-uncased for**

sentiment analysis, both fine-tuned for the task of sentiment classification. The evaluation focused on comparing the performance of these models on our financial news sentiment dataset.

Experimental Setup:

- **Training Procedure:** Both models were fine-tuned for sentiment analysis using the dataset, which included three sentiment labels (positive, neutral, negative). The training process spanned **4 epochs** for both models.
- **Metrics Used:** We evaluated the models using the macro-averaged F1-score to account for class balance and the loss value to assess optimization during training.

Results:

- **DeBERTa-v3:** Achieved a macro-averaged F1-score of **0.856** with a corresponding loss of **0.459**.
- **BERT-uncased:** Achieved a lower macro-averaged F1-score of **0.776** and a higher loss of **0.674**.

Findings: The superior performance of DeBERTa-v3 can be attributed to its pre-training on financial news texts, which better aligns with the domain-specific nature of our dataset. In contrast, BERT-uncased was pre-trained on a broader, more generic corpus, making it less suited to the nuanced language of financial news. This highlights the importance of domain adaptation in achieving higher performance for sentiment analysis tasks in specialized areas.

Overall, the results demonstrate that models pre-trained on domain-specific data, such as DeBERTa-v3, are more effective for sentiment analysis in financial news, both in terms of predictive accuracy and optimization metrics.

5.2 Evaluation of Time Series Forecasting with Sentiment Analysis

As part of this project, we investigated the impact of incorporating news sentiment as an additional feature in time series forecasting models. Specifically, we used XGBoost and Bi-LSTM models to predict daily stock prices, comparing their performance with and without sentiment features.

The inclusion of sentiment scores as features improved the RMSE of the models compared to using financial indicators alone. Specifically:

- XGBoost with sentiment features achieved an average RMSE reduction of **6%** across multiple stocks.
- Bi-LSTM demonstrated an even greater improvement, with average RMSE reductions of **12%**, highlighting its ability to capture both temporal dependencies and sentiment-driven market dynamics.

Findings: The improved RMSE values indicate that incorporating news sentiment provides valuable context for understanding market behavior. Sentiment analysis effectively captured the influence of qualitative factors, such as investor sentiment reflected in financial news, which is often missed by technical indicators alone. These results demonstrate the potential of integrating sentiment analysis for more accurate and interpretable forecasting, particularly in volatile financial environments.

6 Conclusions

At the end of our project, we implemented a visualization component that allows users to see which specific words contributed to the model's predictions. This feature enhances transparency and interpretability, enabling real-world customers to better understand the underlying factors driving sentiment predictions. Such visualization has potential applications in consulting, financial forecasting, and decision-making by providing informative insights rather than opaque, black-box predictions.

Furthermore, our findings demonstrate that augmenting time series forecasting models with sentiment scores leads to better predictive accuracy. By incorporating qualitative factors such as investor sentiment, the models achieve a deeper understanding of market dynamics, enabling more reliable stock price predictions.

This work can be further advanced by incorporating analysis at the level of n-grams instead of single words, allowing the model to capture more complex contextual relationships. Expanding this approach could enhance its applicability to a broader range of use cases, including multi-faceted financial analyses. Additionally, future efforts could focus on integrating this system into real-time analysis pipelines, enabling businesses to process financial news and sentiment in a dynamic, time-sensitive manner. With further refinements, this project can serve as a powerful tool for stakeholders in finance, consulting, and beyond, offering actionable insights based on interpretable and robust sentiment analysis.

7 References

[1] Chan S W K, Chong M W C. Sentiment analysis in financial texts[J]. Decision Support Systems, 2017, 94: 53-64.

[2] Peivandizadeh A, Hatami S, Nakhjavani A, et al. Stock market prediction with transductive long short-term memory and social media sentiment analysis[J]. IEEE Access, 2024.

[3] Mohan S, Mullapudi S, Sammeta S, et al. Stock price prediction using news sentiment analysis[C]//2019 IEEE fifth international conference on

big data computing service and applications (BigDataService). IEEE, 2019: 205-208.

[4] Sousa M G, Sakiyama K, de Souza Rodrigues L, et al. BERT for stock market sentiment analysis[C]//2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI). IEEE, 2019: 1597-1601.

[5] Khedr A E, Yaseen N. Predicting stock market behavior using data mining technique and news sentiment analysis[J]. International Journal of Intelligent Systems and Applications, 2017, 9(7): 22.

[6] Li X, Xie H, Chen L, et al. News impact on stock price return via sentiment analysis[J]. Knowledge-Based Systems, 2014, 69: 14-23.

[7] He P., Gao J., Chen W. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing //arXiv preprint arXiv:2111.09543. – 2021.

[8] Muhammad I., Rospocher M. On Assessing the Performance of LLMs for Target-Level Sentiment Analysis in Financial News Headlines //Algorithms. – 2025. – . 18. – №. 1. – . 46.

[9] Manuel Romero, v3-ft-financial-news-sentiment-analysis, 2024, https://huggingface.co/mrm8488/deberta-v3-ft-financial-news-sentiment-analysis