Multi-Modal Sentiment Analysis for Predicting Stock Market Trends

Youssef Briki
youssef.briki@umontreal.ca

Yuan Li
yuan.li@umontreal.ca

Xintian Xu
xintian.xu@umontreal.ca

Abstract

Social media and news outlets play a crucial role in shaping public opinion and influencing decision-making. This vast amount of user-generated data is often influencial or related to trends in the finance industry, especially the stock market. This project aims to investigate various deep learning models and LLMs, along with variations of training dataset such as social media posts and news articles, in order to compare their performance on predicting the stock market with sentiment analysis. Such models include state-of-the-art transformer-based architectures such as BERT, FinBERT, and Llama3.2 to enhance sentiment prediction accuracy. In addition, macroeconomic numeric data such as GDP, interest rates, inflation etc., and visual finance data charts will also be leveraged in order to examine whether multi-modal sentiment analysis can affect the accuracy of these models. The ultimate goal is to find a system that can provide useful sentiment insights to forecast stock price or market trend.

1 Introduction

Sentiment analysis has emerged as a critical tool for understanding public emotions and opinions. With the advent of deep learning, NLP has significantly improved sentiment analysis accuracy. In the finance industry, social media platforms such as Twitter and Reddit, along with financial and other general news articles, generate vast amounts of text data that reflect real-time sentiments on various topics. They reflect sentiment-laden narratives, from official sources such as large companies and governments down to individuals, that impact market, policies, and consumer behavior. However, these data are often a mixture of true information and opinionated speculations, raising challenges such as sarcasm detection, context understanding, and domain-specific sentiment variations. This project aims to compare the performance of different deep learning models in stock prediction using sentiment analysis. In addition, the effect of the datasets used in training on this task will also be examined, namely news articles, social media posts, macroeconomic data, financial reports, and a combination of these datasets.

2 Related Work

2.1 FinBERT

Although there exist models that can perform well in general sentiment analysis tasks, they may be more lacking in domain-specific tasks due to specialized vocabulary and data of that domain. In stock prediction, an understanding of finance language is crucial and one of the main challenges for sentiment analysis. Therefore, one of the models that we will utilize is FinBERT Araci (2019), a version of BERT Devlin et al. (2019) pre-trained on financial corpus and fine-tuned for sentiment analysis for finance. It was found that FinBERT performed better than Vanilla BERT and other language models on finance tasks. This paper also discusses the uniqueness of sentiment analysis in finance compared to general sentiment analysis tasks, namely that in finance the goal is to utilize sentiment scores to forecast the market. Therefore, this project draws inspiration from this paper, and will also utilize other pre-trained LLMs and fine-tune them for sentiment analysis stock market prediction.

2.2 Financial Sentiment Analysis: Techniques and Applications

In this paper by Du et al. (2024), a review of studies on techniques and applications of financial sentiment analysis is conducted. It discusses the interaction between financial textual sentiments, investor sentiments, and other market information with the market itself and stock prices. Textual sentiment information can be classified as subject or objective, in which the objective is often official information such as macroeconomic data to which investors later react, providing subjective data. Together, these two types of data provide sentiments that can influence stock market behavior. Various techniques used for financial sentiment analysis are reviewed, including machine learning, deep learning, and pre-trained models. Benchmarks such as PhraseBank, FiQA Task 1, SemEval 2017 Task 5 etc. are used for comparison. Although this paper can provide a more comprehensive comparison on how different techniques and models perform on the described benchmark datasets, it provides an overview and knowledge foundation for which we will leverage and consider in our project.

3 Tasks and Datasets

3.1 Tasks

Our main goal is to predict the stock market based on recent data, accompanied by a general summary of public sentiment analysis and confidence in the current state of the market.

We plan to train our model using daily data from the stock market indexes over the past 20 years, focusing on major U.S. indexes and other countries, such as Canada. We are also incorporating macroeconomic data to improve the accuracy of our predictions.

For sentiment analysis, we intend to use both official sources, such as media outlets, and unofficial sources, such as Reddit and stock market subreddits. These platforms gather input from both professional and amateur investors. Analyzing their sentiment and market trust will provide valuable insights to enhance our predictions.

3.2 Datasets

- **Social Media:** Reddit (r/stockMarket, r/stocks, r/StocksAndTraining, etc.)
- **News outlets:** Reuters, Yahoo Finance, CNBC, MoneySense, Investor's Business Daily, etc.
- **Historic stock price:** NASDAQ, NYSE, S&P500 , (2000 - 2022) Toronto Stock Exchange statistics (1956 - present).
- **Macroeconomic:** FRED, World Bank, OECD.

4 Methodology

4.1 Experiment 1: FinBERT & LSTM

For our experiments, aside from the baseline, we will focus on testing multi-modal models. First, we will use FinBERT for generating sentiment score or segmentation for the text data such as news articles and social media posts. We will also use LSTM for analyzing historical stock prices and forecasting future prices. For this part, we will merge stock prices data with macroeconomic data by aligning them on the date, and we will examine whether the addition of macroeconomic data has an effect on price prediction. In order to get our final stock prediction, based on both sentiment analysis and time-series analysis, we will feed the concatenated output of FinBERT and LSTM into a dense network or another type of neural network that remains to be decided, and get the final stock price prediction.

4.2 Experiment 2: Llama 3.2 & LSTM

For our second experiment, we plan on using a LLM, specifically Llama 3.2 (Dubey et al., 2024), for sentiment analysis. Then we will also use LSTM to analyze historical stock prices and macroeconomic data, and combine the outputs of both models to get the final prediction from another neural network. In addition, since Llama 3.2 supports multi-modal data, such as processing of visual data, we will explore whether financial charts can further improve the accuracy of sentiment scores. We picked a LLM such as Llama 3.2 since it may be more capable at handling multi-modal data. Also, since LLMs are trained with much more parameters than BERT or FinBERT, it may have be able to perform better on more tasks and have more contextual understanding. However, we may consider fine-tuning Llama 3.2 on finance data before using it for our experiment, so that it will be more suited for our project.

4.3 Interactive LLM Agent

We will complete this section if time permits and it proves feasible. After running the proposed experiments and identifying the best-performing model, we plan to integrate our tools into an LLM orchestrator. This will create an agentic LLM architecture with a workflow orchestrator, enabling users to ask natural language queries for stock predictions and receive the model's output in a more readable text format.

5 Baselines and Evaluations

5.1 Baselines

To effectively assess the performance of our proposed models, we will compare them against the following baseline approaches:

In the paper byJain and Agrawal. (2024), the authors employ a combination of FinBERT and GAN to predict stock prices, comparing their results with multiple models. Their experimental outcomes will serve as our baseline. Additionally, we will compare the results of applying the model on text-only data and combine text with visual clues. Traditional Financial Indicators: Baseline predictions based purely on macroeconomic indicators (e.g., GDP, inflation, interest rates) without leveraging sentiment data.

5.2 Evaluations

We will evaluate the performance of our models using the following metrics:

Sentiment Classification Metrics:
Accuracy, Precision, Recall, and F1-score for sentiment classification tasks.
AUC-ROC for assessing model discrimination ability.

Stock Market Prediction Metrics:
Directional Accuracy: Measures how often the model correctly predicts the direction of the market movement (up / down).
Mean Absolute Error (MAE) , Root Mean square Error (RMSE): Evaluate prediction errors in stock price forecasting.
Correlation Coefficient (R): Measures the strength of correlation between sentiment-based predictions and actual stock market movements.

Ablation Study: We will conduct ablation studies to assess the impact of different input modalities (text-only, macroeconomic data, and visual clues) on the model's performance.

References

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

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. pages 4171–4186.

Kelvin Du, Frank Xing, Rui Mao, and Erik Cambria. 2024. Financial sentiment analysis: Techniques and applications. *ACM Comput. Surv.*, 56(9).

Abhimanyu Dubey, Abhinav Jauhri, and Abhinav et al. Pandey. 2024. The llama 3 herd of models. *Preprint*, arXiv:2407.21783.

Jainendra Kumar Jain and Ruchit Agrawal. 2024. Fb-gan: A novel neural sentiment-enhanced model for stock price prediction. *Preprint*, arXiv:2407.21783.