ESG and Public Sentiment Analysis for Predicting Short-Term Stock Movement

Tristan Grubbs
Department of Computer Science
University of Virginia School of Engineering and Applied Sciences
Charlottesville, VA, USA
Email: byj4jn@virginia.edu
Computing ID: byj4jn

***Abstract*— This project explores using AI to predict next-day stock price movements from environmental, social, and governance (ESG) sentiment from financial news. Using FinBERT to extract sentiment from headlines, alongside market data, to produce Logistic Regression and XGBoost models. Results show limited predictability with high levels of uncertainty, which underscores the challenge of short-term stock price prediction and the difficulty in analyzing ESG sentiment in the market.**

Keywords*— *ESG, sentiment analysis, FinBERT, stock movement prediction, text mining, logistic regression, XGBoost

I. Introduction

A. Problem

As information has become more readily available to the public the stock markets are increasingly driven by Environmental, Social, and Governance (ESG) narratives and the public's opinion. However, accurate and up-to-date ESG insights are essential to produce an informed investment.

B. Goal

The goal of this project is to develop an AI system that converts financial and ESG news into sentiment and predicts next-day stock movement, whether positive or negative.

II. Importance and Novelty

A. Importance

Public influence on financial markets has grown significantly over time. Individual investors, referred to as *retail investors*, can drive industry investing movements. Popular examples include "meme stocks" such as GameStop ($GME) and AMC Entertainment ($AMC), as well as retail trader darling stocks like Tesla ($TSLA) and ($PLTR). These shifts often occur based on the collective's sentiment rather than traditional fundamentals. An AI system capable of capturing and translating these opinions into investment signals can provide valuable insights for short-term market forecasting.

Beyond retail investors, ESG factors have become increasingly influential for institutional investors. According to *Insights by Stanford Business*, nearly half of respondents report that ESG criteria play a critical role in their decision-making process.

As for why AI-based solutions are best suited for this. Human analysts cannot process thousands of headlines, multiple sentiment signals, and real-time market data all at the same time. AI systems, however, excel in this big data environment and would be able to process large amounts of text and provide rapid evaluation, making them the key to pricing ESG data into the market.

B. Novelty

Most price prediction models use quantitative data from the market, not qualitative data from textual sources. This quantitative data includes recent, historical, and moving averages. Even when looking at AI systems for price prediction, most focus only on numerical data. My system designs a new layer to create an AI system that quantifies how the public and media feel about a company.

III. Data Acquisition and Exploration

This project uses three primary data sources. First, news headlines are collected from Yahoo Finance for each ticker, using the yfinance API, and are supplemented with ESG-focused headlines from the Kaggle ESG News dataset (Table I). Secondly, market data is obtained from the yfinance API which provides price context around the stocks through time. Third, sentiment training data comes from the Financial PhraseBank, a dataset of labeled financial sentences from HuggingFace that verifies accurate sentiment labeling (Table II).

Table I

ESG-focused headlines from the Kaggle dataset.

	sentence	label	positive	negative	neutral	label
1755	The contract …	1	0.039774	0.017756	0.942470	neutral

1281	Kemira shares closed …	1	0.022678	0.105762	0.871560	neutral
350	The company slipped...	0	0.007680	0.974256	0.018065	negative
420	According to …	2	0.961576	0.017403	0.021021	positive

Table II

Financial Phrasebank from HuggingFace.

	sentence	label
0	According to Gran , the company…	1
1	For the last quarter of 2010 …	2
2	In the third quarter of 2010 , net sales …	2
3	Operating profit rose to EUR 13.1 …	2
4	Operating profit totalled EUR 21.1 …	2

IV. Sentiment Scoring Methodology

The sentiment analysis process is built around the FinBERT model, specifically the ProsusAI/finbert implementation, which is pretrained on financial text to associate text with a sentiment score. Before applying the model, I separately validated it with the Financial PhraseBank dataset to ensure accuracy in financial sentiment classification.

For feature engineering, each news headline is scored with positive, neutral, and negative probabilities. These per-headline scores are then aggregated daily (Table III) to create trading signals to be cross-referenced with price data. A sentiment metric is calculated as Positive – Negative, with neutral sentiment being a net zero. This provides a numerical measure of directional sentiment. Additional features, such as sentiment momentum, one, three, and five-day averages, are generated later and used in final predictions with the sentiment scoring. Outputs from the sentiment scoring stage specifically include a table of per-headline sentiment probabilities and the aggregate daily sentiment; these serve as the foundation for integrating sentiment into the predictive models.

Table III.

Aggregated sentiment probabilities daily.

	trading_date	positive	negative	neutral	Sentiment_score
0	2021-10-21	0.013181	0.966259	0.020559	-0.953078
1	2021-10-24	0.501397	0.027248	0.471354	0.474149
2	2021-10-25	0.077350	0.537838	0.384813	-0.460488
3	2021-10-26	0.248237	0.670894	0.080869	-0.422657
4	2021-10-28	0.150390	0.259553	0.590058	-0.109163

V. Predictive Modeling

The prediction task is formulated as a binary classification problem where the goal is to predict whether a stock will close higher, indicated as 1, or lower, indicated as 2, on the next trading day. The feature set includes the sentiment data sentiment_score and sentiment_momentum, along with short-term return data return_1d, return_3d, and return_7d.

I used the data I prepared to train two models. Logistic Regression is the basic model due to its simplicity and provides a baseline. On the other hand, the XGBoost model uses a gradient-boosted decision tree algorithm and is better at capturing the nonlinear interactions between sentiment and price momentum. Depending on the stock ticker, either model may ultimately outperform the other.

Model performance is evaluated using accuracy and F1 score as primary metrics. Additionally, SHAP values are applied to demonstrate feature importance, highlighting the extent to which sentiment or price-based features drive predictions (Figure I.).

For example, in this report, I ran my models on the NVIDIA ticker ($NVDA). Results showed that Logistic Regression achieved an accuracy of 0.4312 with an F1 score of 0.4681, while the XGBoost model performed better with an accuracy of 0.5455 and an F1 score of 0.5238 (Figure II.). These results show modest prediction power with XGBoost, particularly providing a slight improvement over the baseline.

Figure I.

Graph showing multiple features implemented into the model, graphed with SHAP values on the x-axis and Feature Value on the y-axis.

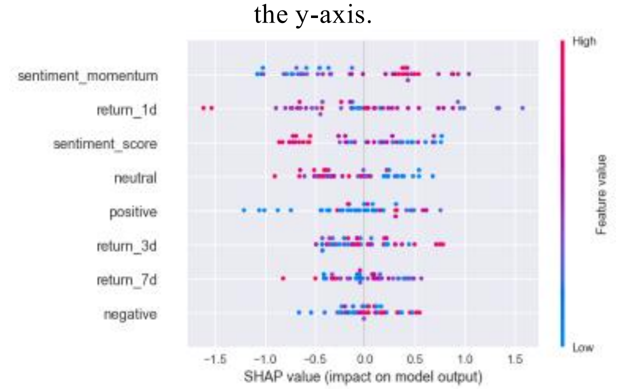
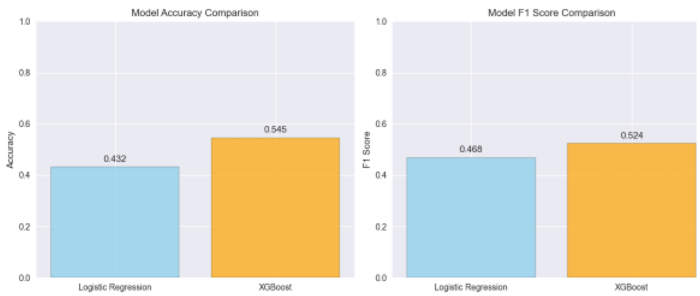


Figure II.

Simple bar graph showing how the Logistic Regression and XGBoost accuracy and F1 values are reported by the system.



VI. Real-Time Prediction System

The real-time prediction system is the last implemented section of the system and is designed to retrieve live data, process sentiment, and generate predictions using pre-trained models. The pipeline begins by pulling the last 5 days of headlines through the yfinance API and retrieving the last month of daily price history. The price history is used to calculate the momentum features.

Feature engineering happens in the same section as the real-time predictions. The recent headlines that were pulled with yfinance are scored using FinBERT for positive, negative, and neutral probabilities, which are then aggregated into a daily sentiment_score, just like how data was prepared when training the AI. Additional features include sentiment_momentum, which measures how sentiment has changed compared to the previous day, and sentiment_avg_5d, which is a 5-day average to smooth outlying noise. Beyond the sentiment features, price-based features like return_1d, return_3d, and return_7d capture short-term market trends that are the traditional basis of stock prediction. All of these are presented with each AI's output (Figure III).

Inference logic combines outputs from both models. Logistic Regression is a standard prediction of the probability of upward movement. XGBoost applies a heuristic adjustment to its raw probability using the formula:

Logit_adj = Logit_raw + (1.2 × Score + 0.8 × Momentum + 0.5 × Avg_5d).

This adjustment amplifies the signals when strong sentiment or momentum. The prediction is then checked to avoid extreme confidence and clipped between 0.05 and 0.95.

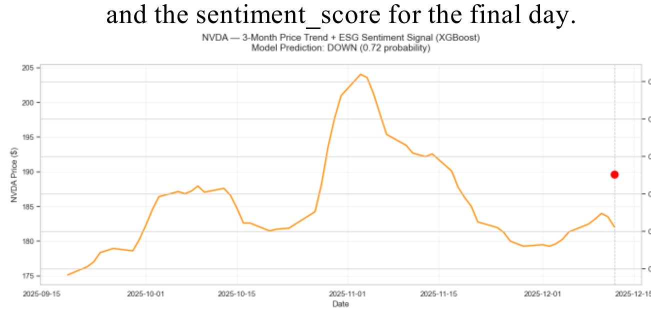
Finally, the system outputs the predicted direction (UP or DOWN), along with the raw and adjusted probabilities to promote transparent AI. A contextual chart is then generated, showing the last three months of price history and the final numerical sentiment probability marked on the final day (Figure IV).

Figure III.
XGBoost AI output after reporting all feature values and final prediction.

```
[XGBoost] Ticker: NVDA
Lookback: last 5 day(s)
Headlines used: 10

Features:
  positive: 0.6873
  negative: 0.1521
  neutral: 0.1605
  sentiment_score: 0.5352
  sentiment_momentum: 0.5215
  return_1d: -0.0327
  return_3d: -0.0538
  return_7d: -0.0254
Model: XGBoost | Prediction: DOWN | P(UP) raw = 0.117 | adjusted = 0.277
```

Figure IV.
XGBoost chart showing the last three months of price history and the sentiment_score for the final day.



VII. Results Analysis and Discussion

Both Logistic Regression and XGBoost achieved F1 scores in the range of 0.40-0.50, which is only slightly better than if you were to randomly guess. Logistic Regression demonstrated more stable guessing behavior and higher recall. XGBoost captured nonlinear feature interactions but exhibited higher variance. Due to the small time window, XGBoost produced some unstable probability estimates. Depending on the historical nature of the stock price and the effects ESG has had on it, either model could be better suited for predictions.

Short-term stock movement is extremely noisy and is affected by macroeconomic movements, earnings releases associated with the stock or related stocks, and market-wide momentum that is not always apparent from the headlines. Positive sentiment does not consistently imply upward price movement. The models learned to weigh recent price momentum against sentiment rather than relying on sentiment alone, which would often fail to produce accurate results.

Several challenges were encountered during modeling. ESG headlines provided a weak signal strength to stock movement because they reflect more on the long-term fundamentals of a stock, rather than next-day price changes.

Furthermore, a small sample size available for training reduced the effectiveness of the complex XGBoost model. Probability miscalibration was the last major issue I faced. Models, particularly XGBoost, frequently produce overconfident predictions that are not accurate to the model's prediction ability.

To mitigate these issues, 5-day sentiment averages and sentiment momentum features were implemented. Probability caps and floors, as well as heuristic adjustments, were applied to reduce overconfidence. F1 scores were given special emphasis, and SHAP explanations are provided to maintain transparency throughout the system.

The key lesson learned is that sentiment-based prediction alone is insufficient for generating high-accuracy trading signals. However, sentiment features improve contextual understanding of the market, provide insights into stock changes over time, and are best used as decision support tools. Based on this, I think the best course of action would be to use sentiment-based features as part of a much larger whole trained on as much data as possible. The sentiment-based features should be weighed more heavily in long-term stock trends rather than short-term. The sentiment features' qualitative data and the quantitative data in traditional stock prediction systems would be combined to create a powerful system to assist institutional investors or retail investors, although retail investors would likely need a less powerful, light-weight system for home use.

VIII. Conclusion

A. Summary of Work

I developed a system to predict next-day stock movements using ESG-focused news sentiment and FinBERT. This system integrates historical data analysis, automated sentiment scoring, and the machine learning models Logistic Regression and XGBoost, all into a system of notebooks that house different sections of the process

B. Key Findings

ESG sentiment alone is a weak predictor of daily price action, standing at between 45-55% accuracy. This confirms that news is often mostly priced into the market by the time it reaches the public, and beyond that reflects shifting fundamentals that take long periods of time to significantly affect the stock price.

Due to the limited datasets and non-linear relationships of features used in the models, Logistic Regression and XGBoost have their own strengths and weaknesses, and which is better for prediction depends on the particular stock being predicted.

While not being the best standalone trade signal, the sentiment features do provide useful context are should be implemented in stock prediction systems moving forward.

C. Future Changes

In the future, I would expand the dataset to include more general financial news and headlines from other sources. Particularly, Reddit was a source I was not able to capture during this project, but it would be nice to capture, as it is the host to message boards used by a significant number of retail investors that can have a large impact on the market. Lastly, I would also shift the predictions to be for the next week or month with a rolling prediction model. The rolling model would accumulate ESG sentiment and adjust predictions in real time to fit the changing context around a company's stock price. This would be better suited for the long-term nature of ESG.

References

[1] OpenAI. (2025). *ChatGPT (GPT-5.2, Dec 2025 version)* [Large language model]. https://chat.openai.com/

Statement of use: I used ChatGPT to plan and refine my AI Final Project Report. Specifically to plan out what sections would be required to wholly explain my project and answer the questions provided

[2] Anthropic. (2025). *Claude Sonnet 4.5* [Large language model]. https://www.anthropic.com/
Statement of use: I used Claude Sonnet 4.5 to assist with coding, debugging, and refinement of machine learning systems. Claude was primarily used to help integrate sentiment features, adjust model behavior, and resolve implementation issues during development.

[3] Microsoft. (2025). *Word Copilot* [Writing Assistent]. https://github.com/features/copilot
Statement of use: I used Word Copilot to help wth my writing and recommend the layout of additions in my report.

[4] Google. (2025). *Gemini 3 Pro* [Large language model]. https://deepmind.google/technologies/gemini/
Statement of use: I used Gemini 3 Pro for conceptual clarification and alternative explanations of machine learning models and evaluation metrics. I also used it as an alternative coding tool, although Claude Sonnet 4.5 was primarily used for that.

[5] Takala, P. (n.d.). *Financial PhraseBank*. Hugging Face Datasets. https://huggingface.co/datasets/takala/financial_phrasebank

[6] Debashish. (2022, November 30). *Daily ESG news and headlines*. Kaggle. https://www.kaggle.com/datasets/debashish311601/daily-esg-news-and-headlines

[7] Gelfand, A., Larcker, D. F., Lee, L.-E., Seru, A., & Tayan, B. (2022, November 10). *Big investors say they use ESG to reduce risk (but mostly focus on the E and G)*. Stanford Graduate School of Business. https://www.gsb.stanford.edu/insights/big-investors-say-they-use-esg-reduce-risk-mostly-focus-e-g

[8] Zucchi, K. (n.d.). *Institutional vs. retail traders: Key differences explained*. Investopedia. https://www.investopedia.com/articles/active-trading/030515/what-difference-between-institutional-traders-and-retail-traders.asp

[9] Wikimedia Foundation. (2025, November 16). *Meme stock*. Wikipedia. https://en.wikipedia.org/wiki/Meme_stock

[10] Staff. (2021, April 21). *The 3 darlings of WallStreetBets: What traders need to know*. StocksToTrade. https://stockstotrade.com/3-darlings-of-wallstreetbets/