LLM-Based Commodity News Sentiment Analyzer for Market Stabilization

Basar Temur, 1909058,
basar.temur@students.uni-mannheim.de

March 29, 2025

1 Introduction

Market dynamics and price volatility constitute fundamental challenges in commodity markets, particularly for essential goods such as food and energy. These fluctuations are often intensified during periods of macroeconomic uncertainty, climate events, or geopolitical disruptions. Such shocks can severely undermine both producers and consumers: producers may experience unsustainable financial losses during critical phases such as harvest, while consumers may face inaccessibly high prices. A robust institutional mechanism is necessary to stabilize these markets. To this end, state-sponsored or producer-led entities—such as Commodity Marketing Boards (CMBs) and State Trading Enterprises (STEs)—have historically played a vital role in safeguarding market functionality. These organizations provide risk-sharing mechanisms analogous to futures contracts, enabling producers to secure their margins by locking in prices in advance. Additionally, these entities are well-positioned to leverage advanced technologies, real-time market intelligence, and regulatory capital buffers (sometimes backed by sovereign guarantees) to absorb exogenous shocks. Notably, such organizations can maintain capital reserves—often referred to as capital buffers or stabilization reserves—which serve as equity cushions against adverse market developments [2].

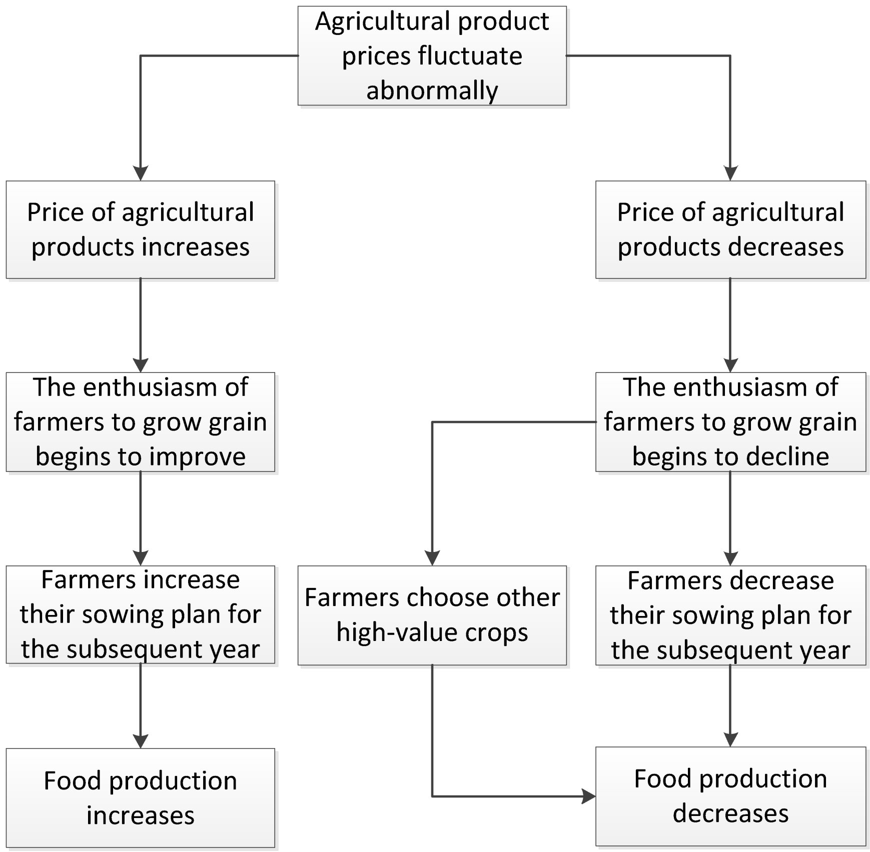

Figure 1: Diagram showing the effect of price fluctuations on farmer's behaviour [7]

Within this project a LLM(FinBert) based commodity news sentiment analyzer was developed, possibly for such an organization described above. Last minute financial news are most of the time the main driver of price fluctuations and are a major factor for traders when making decisions. Sometimes their effect may be factored in instantly and sometimes over a longer period. It is also possible they may not have an impact on current trends at all too. Therefore it is important to make predictions over a future time horizon rather than current conditions. In this project in order to smooth out noise in financial data and determine trading signals, crossovers between long term and short term moving average was used. For many traders short term moving average rising above long term moving average may be interpreted as a trading signal. However because it is based on retrospective data it is important to predict those signals beforehand. In this project what is being predicted is possible change in trading signals in a future 5 day time window caused by last-minute news.

2 Data

The dataset utilized in this study is derived from the Kaggle project "Corn and Soybean 2014–2020" [10]. It encompasses news articles and headline data related to corn and soybean from soybeansandcorn.com. Daily closing prices were extracted from the Chicago

Board of Trade (CBOT), accessed via Macrotrends. While the current implementation is constrained to daily news feeds, several authoritative market reports—such as the USDA's WASDE (World Agricultural Supply and Demand Estimates)—present significant potential for augmenting predictive power in future iterations.

Source	**Type**	**Time Range**
soybeansandcorn.com	News Headlines	2014–2020
Macrotrends	CBOT Close Prices	2014–2020
USDA WASDE	Market Reports	Not used (future)

Table 1: Summary of data sources.

3 Preprocessing and Parameter Decisions

The selection of time windows for calculating moving averages was empirically determined based on the frequency and clarity of trading signals. Extended time windows (e.g., 30-day vs. 7-day) tend to over-smooth the data, thus suppressing short-term volatility and potential signal generation. Therefore, a compromise was struck by selecting a 5-day short-term window and a 20-day long-term window. The prediction horizon was also set at 5 days, consistent with the short-term signal propagation timeframe observed in financial markets by Lo and MacKinlay [6]. The model interprets short-term moving averages rising above the long-term average as a bullish signal and the inverse as bearish. Neutral signals indicate no significant trend deviation. However, due to the retrospective nature of moving averages, the optimal trading window lies between the news event and the crossover—a period where conventional market actors are yet to react. This model is designed to capitalize on that latency.

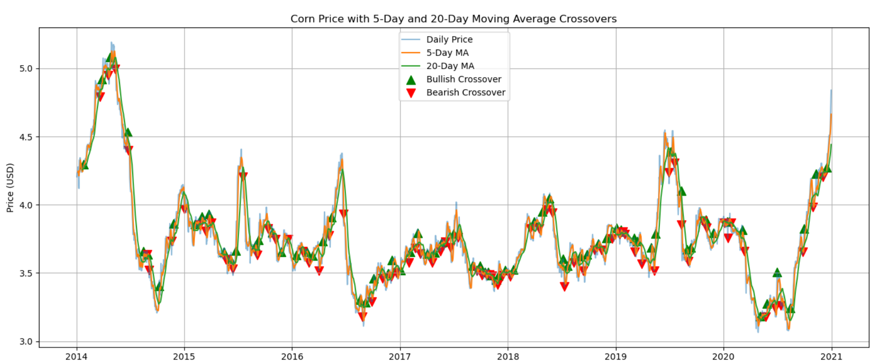

Figure 2: Overlay of short-term and long-term moving averages with crossover signals.

4 LLM Model Selection

The selected NLP model for this study was FinBERT from ProsusAI [9], a fine-tuned variant of BERT specifically adapted for financial sentiment analysis. Pretrained on a domainspecific corpus of financial texts, FinBERT offers enhanced accuracy in classifying financial sentiment compared to generic BERT variants [11]. Initially, full news articles were considered for sentiment classification. However, BERT-based models impose a 512-token input limit, whereas many articles exceed 700–800 tokens. While truncation and padding are available, truncation risks discarding salient information. Consequently, headlines were used as proxies, given their compact and information-dense nature. For future enhancements, neural network-based summarization (e.g., BART, T5) or transformer architectures supporting longer context windows (e.g., Longformer, BigBird) may be considered [3].

Model	**Token Limit**	**Specialization**
BERT	512	General NLP
FinBERT	512	Financial Texts
Longformer	4096+	Long Documents
BigBird	4096+	Sparse Attention

Table 2: Comparison of transformer models.

5 Coding and Training

Model development was conducted with assistance from ChatGPT-4o [8]. The training process encountered class imbalance, a well-known limitation in financial sentiment datasets where neutral signals dominate. While oversampling/undersampling were considered, the approach employed here was class weighting, a strategy that modifies the loss function to penalize misclassification of minority classes more heavily. A custom loss function—Focal Loss—was implemented using Hugging Face's Trainer subclassing method, following guidelines proposed by Alanturner2 [1]. This approach emphasizes harder-to-classify examples by reducing the relative loss for well-classified instances [5].

Focal Loss was used to address imbalance:

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t) \tag{1}$$

6 Evaluation and Performance

Despite applying class weighting, the model exhibited weak performance, particularly for the minority bullish and bearish classes. As is typical in financial sentiment tasks, the classifier showed an overfitting bias toward the neutral class, which dominated the training set. The evaluation metrics are summarized below:

Class	**Precision**	**Recall**	**F1-Score**	**Support**
-1	0.094	0.143	0.113	21
0	0.817	0.686	0.746	156
1	0.062	0.111	0.080	18

Table 3: Classification Report

	-1	0	1
-1	3	12	6
0	25	107	24
1	4	12	2

Table 4: Confusion Matrix

These metrics reflect the common challenge in financial text classification: accurate modeling of rare yet high-impact signals. Future efforts may benefit from ensemble methods or synthetic minority oversampling (SMOTE) to improve class balance [4].

7 Next Steps and Conclusion

For subsequent research, several pathways can enhance model fidelity:

1. **Expanded Dataset:** Increasing the temporal scope of training data would improve model generalization and allow the model to learn from a broader array of market cycles and news sentiments.

2. **Token Limit Resolution:** Employing advanced summarization architectures (e.g., BART, T5) or extended-context transformers (e.g., Longformer, BigBird) could enable full-text processing while preserving crucial information lost during truncation.

3. **Data Balancing Strategies:** Exploring hybrid methods such as a combination of SMOTE and class-weighted loss functions may mitigate the impact of class imbalance on model learning.

4. **Backtesting:** Simulating the model's predictions in historical trading environments would provide a more accurate assessment of its real-world financial viability.

5. **Integration of WASDE Reports:** Structured insights from USDA's WASDE market reports could complement unstructured news data, creating a multimodal input strategy suitable for ensemble models.

In its current form, the study offers a compelling proof-of-concept that highlights both the promise and limitations of leveraging large language models for commodity price prediction. Although the model is not yet suitable for deployment in algorithmic trading, it lays a solid foundation for future enhancements.

The approach demonstrates how sentiment-aware forecasting could be integrated into the operational frameworks of state-backed commodity stabilization mechanisms, thereby contributing to more resilient and informed agricultural policy and trade decisions.

References

[1] Alanturner2. Create a weighted loss function to handle imbalance, 2023. https://discuss.huggingface.co/t/create-a-weighted-loss-function-to-handle-imbalance/138178/3.

[2] Basel Committee on Banking Supervision. Basel iii: A global regulatory framework for more resilient banks and banking systems, 2011.

[3] Iz Beltagy, Matthew E Peters, and Arman Cohan. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*, 2020.

[4] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[5] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988, 2017.

[6] Andrew W Lo and A Craig MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

[7] Abdelbaset M. Nassani, Abdullah M. Aldakhil, Qaisar M. Abro, Khalid Zaman, and Boqiang Lin. Environmental challenges and green economy: The role of the renewable energy consumption, co emissions, and trade openness in sustainable development. *Sustainability*, 9(6):906, 2017.

[8] OpenAI. Chatgpt-4o, 2024. https://chat.openai.com.

[9] ProsusAI. Finbert, 2020. https://huggingface.co/ProsusAI/finbert.

[10] Ivan F Reis. Corn and soybean 2014–2020, 2023. https://www.kaggle.com/datasets/ivanfilhoreis/corn-and-soybean-news-and-trend.

[11] Yi Yang, Chen Sun, Yuan Tian, and Hong Wu. Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*, 2020.

A Appendix

Jupyter Notebook

B Appendix

ChatGPT History