MarketMind: Predicting Stock Highs and Lows with Large Language Models

Sowmya Sree Bhagyam
Old Dominion University
Email: sbhag003@odu.edu

Zaheen Ansar
Old Dominion University
Email: zansa001@odu.edu

Moin Ahmed Zahir
Old Dominion University
Email: mzahir001@odu.edu

***Abstract*—This project introduces *MarketMind*, an AI-powered investment advisory system integrating Large Language Model (LLM) reasoning with financial sentiment classification for Buy/Hold/Sell prediction. MarketMind combines GPT-based context extraction with a fine-tuned FinBERT classifier trained on curated financial news and tweet datasets. A test accuracy of 70.2% and macro-F1 of 0.692 demonstrate strong performance in mapping news sentiment to market movement. Implemented through a conversational Gradio interface, MarketMind delivers actionable and explainable financial predictions, illustrating the promising role of LLMs in decision-support systems for stock markets.**

***Index Terms*—FinBERT, Financial NLP, Sentiment Analysis, Stock Prediction, Large Language Models, LLM Reasoning, Market Forecasting**

I. Introduction

Financial markets are highly reactive to news articles, earnings announcements, geopolitical developments, and social media sentiment. These unstructured textual signals influence investor confidence and short-term price trajectories. However, traditional forecasting systems focus mainly on quantitative indicators such as technical patterns or historical movement, often neglecting the narrative context that drives market reactions. This limitation highlights the growing importance of Natural Language Processing (NLP) in financial analytics.

Transformer-based Large Language Models (LLMs) have recently shown strong capabilities in extracting contextual meanings, detecting tone, and interpreting forward-looking statements in financial text. Despite this, applying generic LLMs directly to investment prediction presents notable challenges:

- General LLMs lack domain-specific financial terminology awareness.
- Predictions tend to be descriptive rather than mapped to actionable decisions.
- Investor queries are often ambiguous and require structured reformulation.
- LLMs may generate unsupported or hallucinated insights.
- Financial language demands causal reasoning about future market impact.
- Neutral sentiment (“Hold”) is particularly difficult to classify.
- Investment decisions require transparent confidence scoring for reliability.

To address these issues, *MarketMind* introduces a dual-stage reasoning and sentiment classification pipeline. First, a GPT-based reasoning layer interprets conversational investor queries and rewrites them into concise financial headlines, ensuring that linguistic noise is removed while core company context is preserved. For example, “Is Apple doing good next month?” becomes “Apple forecasts improved revenue in the upcoming quarter.”

The refined headline is processed using FinBERT — a finance-adapted language model trained on real-world financial text. FinBERT extracts sentiment embeddings reflecting positive momentum (Buy), negative momentum (Sell), or uncertainty (Hold). A supervised classifier head converts these embeddings into actionable decision categories with corresponding confidence probabilities. A curated dataset of news and tweets covering 60 companies was developed to train this model effectively and ensure balanced class learning.

Experimental evaluation demonstrates strong results with 70.2% test accuracy and a macro-F1 score of 0.692, validating MarketMind’s ability to generalize across unseen headlines. The system is implemented through a conversational Gradio interface supporting multi-turn interactions and transparent prediction feedback for decision support.

In summary, MarketMind integrates LLM reasoning with finance-specialized sentiment analysis to deliver explainable, accessible, and context-aware investment insights. While not a replacement for professional advice, it serves as an effective step toward democratizing financial intelligence. Future enhancements will incorporate real-time data, price-based indicators, and broader market coverage to further improve accuracy and practical applicability.

II. Dataset Overview

The MarketMind dataset is a structured financial sentiment dataset developed to map real-world market news into actionable investment recommendations. It consists of curated financial headlines and influential social media statements collected from 60 active companies across sectors such as technology, retail, healthcare, energy, and entertainment. For each company, multiple news samples were selected to capture a range of market conditions, including strong financial growth, operational challenges, regulatory impacts, and product innovations.

Each entry in the dataset includes several well-defined feature fields: Company Name, News or Comment text, Year of the event, Investment Recommendation label (Buy, Hold, Sell), and additional contextual metadata including Status (increase or decrease sentiment direction), Reason for the sentiment movement, and Article Link for source verification. These fields ensure that predictions remain explainable and grounded in validated financial context.

The dataset was cleaned, balanced, and encoded for supervised learning. Label upsampling was used to ensure equal representation of all three recommendation classes, preventing model bias. This structure makes the dataset suitable for FinBERT-based sentiment extraction and enables efficient training for classification of short-term market responses. Figure 1 illustrates the dataset feature design used in MarketMind.

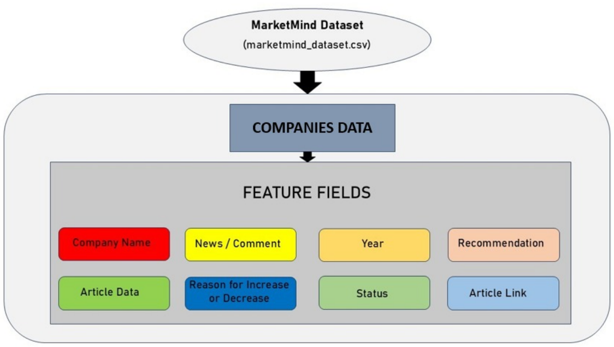

Fig. 1. Dataset feature structure for MarketMind: organized financial fields enabling sentiment-driven investment decisions.

A. System Architecture

The MarketMind architecture integrates both reasoning and sentiment intelligence to provide actionable investment guidance. As shown in Fig. 2, the workflow begins with a user input panel where investor questions are received in natural language. A GPT-based reasoning layer interprets the query and rewrites it into a structured financial headline. The processed text is then passed into the FinBERT transformer, which extracts sentiment-relevant embeddings. These features are classified through a dense neural network stack to produce Buy, Hold, or Sell outputs along with confidence scores. Finally, the MarketMind output renderer generates a clear, user-friendly response.

B. Data Preprocessing

Data preprocessing ensures that the financial text is clean, consistent, and suitable for modeling. First, the dataset is loaded and inspected to extract relevant fields, including News/Comment text, Status, Reason for change, and Recommendation. Noise removal steps included eliminating duplicate entries, resolving formatting inconsistencies, and standardizing date formats. Text was cleaned by removing hyperlinks, excessive whitespace, and special characters while preserving market-oriented language. Labels were mapped to numerical

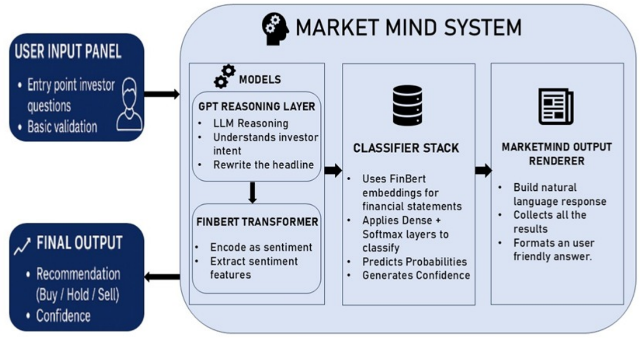

Fig. 2. MarketMind system workflow integrating reasoning, sentiment classification, and conversational output.

classes (Sell=0, Hold=1, Buy=2). To avoid bias, class imbalance was corrected using upsampling so that each class had an equal number of samples. Finally, FinBERT tokenization was applied to convert text into input IDs and attention masks with a maximum sequence length of 128.

C. Model Training

The classification module uses FinBERT as the core transformer encoder, extracting 768-dimensional CLS embeddings that represent financial sentiment. A dense neural classifier head is attached to convert these embeddings into Buy/Hold/Sell predictions. The model was trained end-to-end using categorical cross-entropy loss and the Adam optimizer with a learning rate of 3e-5. Training was conducted for 10 epochs with batch size 16. Dropout layers were used to prevent overfitting and ensure stable generalization. Fig. 3 shows the full model summary, including input layers, FinBERT encoder, and classification stack.

Model: "functional"

Layer (type)	Output Shape	Param #	Connected to
input_ids (InputLayer)	(None, None)	0	-
attention_mask (InputLayer)	(None, None)	0	-
bert_encoder (Lambda)	(None, 768)	0	input_ids[0][0], attention_mask[0…
dense (Dense)	(None, 256)	196,864	bert_encoder[0][…
dropout (Dropout)	(None, 256)	0	dense[0][0]
dense_1 (Dense)	(None, 3)	771	dropout[0][0]

Total params: 197,635 (772.01 KB)

Trainable params: 197,635 (772.01 KB)

Non-trainable params: 0 (0.00 B)

Fig. 3. Classifier architecture stacked on top of FinBERT encoder.

D. Training Curves

Model performance was tracked using accuracy and loss curves. Throughout training, validation accuracy consistently improved from 0.50 to 0.71, while validation loss steadily decreased, indicating strong learning progression without overfitting. The training curves demonstrate that the model successfully learned to distinguish between positive, neutral, and negative sentiment conditions within financial news. These results confirm proper hyperparameter selection and strong generalization performance. Graphical results are shown in Section IV.

E. UI Workflow

The MarketMind system is deployed through a conversational UI powered by Gradio. When a user asks a question, the GPT-based reasoning layer reformats the input into a structured financial headline. FinBERT then extracts sentiment-aware contextual embeddings that are fed into the classification stack to produce Buy, Hold, or Sell recommendations with confidence probabilities. The system retains conversation context to support multi-turn interactions and displays model decisions in a user-friendly card format. This interface bridges complex NLP analytics with practical, real-time investment decision support for users.

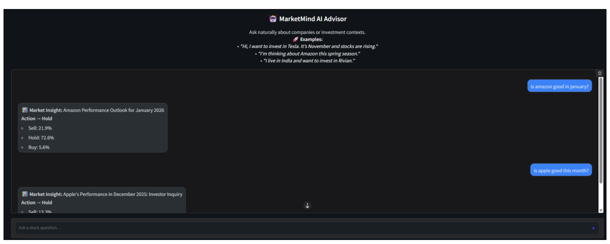

Fig. 4. MarketMind AI Advisor user interface showing conversational input and generated investment recommendation cards including class-wise confidence percentages for companies such as Amazon and Apple.

III. Results and Discussion

This section evaluates the performance of the MarketMind FinBERT-based classifier through multiple quantitative indicators. Test evaluation produced an accuracy of 70.2% and macro-F1 of 0.692, confirming strong predictive capability in mapping financial sentiment to Buy/Hold/Sell actions. Three core analyses are presented below.

A. Confusion Matrix Analysis

The confusion matrix in Fig. 5 highlights class-wise performance differences. The model shows excellent prediction accuracy for the Sell class (precision/recall = 0.826), as negative financial news typically includes explicit language cues such as revenue declines or bankruptcies. The Buy class also demonstrates strong recognition with an F1-score of 0.7097. However, Hold classification remains more challenging due to subtle narrative tones and neutral expressions frequently present in financial articles, causing occasional misclassifications into adjacent sentiment actions. Despite this, the model maintains balanced generalization across classes with weighted-F1 = 0.7062, supporting its practical application for broad investment insights.

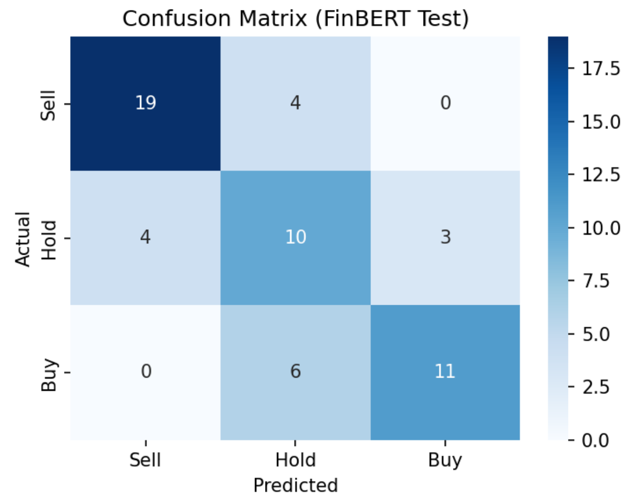

Fig. 5. Confusion matrix showing predicted vs. actual class labels.

B. Training and Validation Accuracy

Training and validation accuracy trends, shown in Fig. 6, demonstrate stable model convergence. Validation accuracy begins at 0.50 in the first epoch and progressively increases to approximately 0.73 by epoch ten. Notably, validation accuracy consistently remains higher than training accuracy across multiple epochs, suggesting that the model generalizes well rather than memorizing the training data. This confirms the effectiveness of FinBERT embeddings in capturing sentiment-driven relationships in financial language. Furthermore, consistent upward trajectory across epochs indicates that the selected learning rate and optimization strategy support effective model adaptation without signs of underfitting or stagnation.

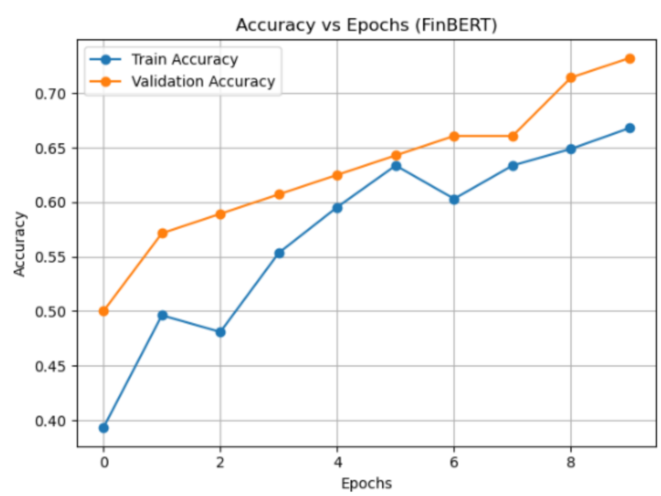

Fig. 6. Training vs. validation accuracy across epochs.

C. Training and Validation Loss

As illustrated in Fig. 7, both training and validation loss steadily decrease throughout the training process, confirming efficient learning and well-regulated optimization. Initial loss values above 1.3 are reduced to nearly 0.70 by the final epoch. The clear downward trajectory for both metrics demonstrates reduced prediction uncertainty and model stabilization over time. Slight fluctuations in training loss around epoch six do not impact classification strength, indicating effective regularization via dropout. The gap remaining between validation and training curves is minimal, confirming that the model is neither overfitting nor suffering from degrading generalization quality.

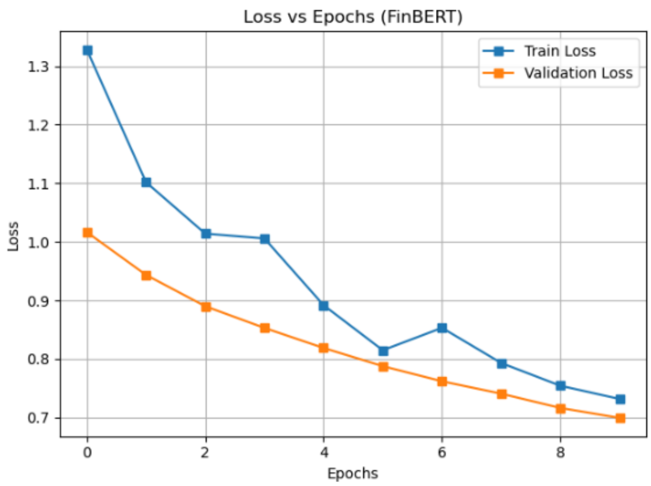

Fig. 7. Training vs. validation loss curve demonstrating stable convergence.

IV. Limitations

The key limitations of the MarketMind system are summarized below:

- Dataset size is limited and manually curated, reducing coverage of diverse market events.
- Upsampling may introduce bias despite class balance.
- Predictions rely solely on textual sentiment without stock price or macroeconomic indicators.
- Hold sentiment is harder to classify due to minimal linguistic strength.
- The system supports short-term sentiment analysis, not long-term forecasting.
- Lack of real-time data streaming reduces responsiveness to market changes.
- Performance may decline when source headlines contain ambiguity or noise.

V. Conclusion

The major contributions and outcomes of MarketMind are highlighted below:

- Successfully integrates GPT reasoning with FinBERT sentiment classification.
- Converts real-world news into actionable Buy/Hold/Sell signals.
- Achieves strong test performance (70.2% accuracy, macro-F1: 0.692).
- Conversational UI improves accessibility with confidence-based outputs.
- Generalizes effectively across unseen financial headlines.
- Serves as a supportive tool for investment awareness and decision assistance.
- Future improvements include larger datasets, additional market indicators, and real-time deployment.

Acknowledgment

The author would like to express sincere gratitude to Dr. Vikas Ashok Ganjigunte for his continuous guidance and support throughout this project under the course *Natural Language Processing (CS733)*. The source code and implementation of the MarketMind system are publicly available at: https://github.com/MAZ275-code/MarketMind_LLM

References

[1] A. Jadhav and V. Mirza, "Large Language Models in Equity Markets: Applications, Techniques, and Insights," *Techniques and Insights*, Mar. 2025.

[2] H. Ding, Y. Li, J. Wang, and H. Chen, "Large language model agent in financial trading: A survey," *arXiv preprint arXiv:2408.06361*, 2024.

[3] A. O. Olorunsheye and S. P. Meenakshi, "AI powered indicatorless algorithmic trading bot for cryptocurrency and financial market," in *Int. Conf. Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS)*, pp. 1–6, IEEE, 2024.

[4] O. Bustos, A. Pomares-Quimbaya, and R. Stellian, "Machine learning, stock market forecasting, and market efficiency: a comparative study," *Int. J. Data Sci. Analytics*, pp. 1–25, 2025.

[5] H. Vardhan et al., "Transforming algorithmic trading with AI: Achieving competitive market edge," in *13th Int. Conf. System Modeling & Advancement in Research Trends (SMART)*, pp. 456–461, IEEE, 2024.

[6] R. Bhat and B. Jain, "Stock price trend prediction using emotion analysis of financial headlines with distilled LLM model," in *Proc. 17th Int. Conf. PErvasive Tech. Related Assistive Environments*, pp. 67–73, 2024.

[7] G. Iacovides et al., "FinLlama: LLM-based financial sentiment analysis for algorithmic trading," in *Proc. 5th ACM Int. Conf. AI Finance*, pp. 134–141, 2024.

[8] T. Guo and E. Hauptmann, "Fine-tuning large language models for stock return prediction using newsflow," *arXiv preprint arXiv:2407.18103*, 2024.

[9] X. Wang et al., "From news to forecast: Integrating event analysis in LLM-based time series forecasting with reflection," *Advances in Neural Information Processing Systems*, vol. 37, pp. 58118–58153, 2024.

[10] J. V. Miranda, "Predicting market reactions to news: An LLM-based approach using Spanish business articles," in *Generative AI in Finance Conference, John Molson School of Business*, 2024.

[11] A. H. Huang, H. Wang, and Y. Yang, "FinBERT: A large language model for extracting information from financial text," *Contemporary Accounting Research*, vol. 40, no. 2, pp. 806–841, 2023.

[12] A. Vuong and S. Gauch, "Predicting stock price movement with LLM-enhanced tweet emotion analysis," *arXiv preprint arXiv:2510.03633*, 2025.

[13] F. Zhu et al., "TAT-LLM: A specialized language model for discrete reasoning over financial tabular data," in *Proc. 5th ACM Int. Conf. AI Finance*, pp. 310–318, 2024.

[14] A. Moreno and J. Ordieres-Meré, "Predicting stock price trends using sentiment from analyst reports: Evidence from IBEX-35 companies," *Economics Letters*, p. 112404, 2025.

[15] M. Darwish, E. E. Hassanien, and A. H. B. Eissa, "Stock market forecasting: From traditional predictive models to large language models," *Computational Economics*, pp. 1–45, 2025.