Text2Cash - M^3 Sentiment Analysis

Ryan Engel
Stony Brook University

Gowhith Gandem
Stony Brook University

Mario Xerri
Stony Brook University

Abstract

We propose a multi-modal framework for forecasting stock returns by combining a news sentiment model with a multivariate time-series foundation model (TSFM). We utilize the FMP Cloud News data source, and a transformer language model to generate contextualized embeddings of news text on each given day. In parallel, we use CRSP price data to construct approximately 25 technical indicators via Pandas-TA (an open-source technical analysis python library), which are input into a multivariate TSFM known as Moirai. This model produces latent embeddings for each of the technical indicator time-series, capturing historical market dynamics up to the given day. A gating mechanism is used to combine the two embedding spaces, and a logistic regression head is subsequently trained to produce a contextualized sentiment score, representing the model's prediction of positive future returns. Finally, we implement and backtest trading strategies based on the model's signals, evaluating their performance and tradeoffs. Our results show strong performance, outperforming simple baselines of the SPY Index and an equal-weight strategy. The results show strong alpha generation, with the strongest model giving a sharpe ratio of 0.95 and cumulative returns of 195.82% over the 5-year backtest.

1 Introduction

Large Language Models (LLMs) have gathered significant attention for their ability to interpret large amounts of data and to make meaningful predictions from such data. However, their applications in finance remain an emerging field. Analysis of financial data text data for predicting stock returns involves converting the text into a vector representation which will be passed to a predictive model. Trained on vast and diverse data, transformers can generate rich, context-aware vector representations of text (Vaswani et al., 2023).

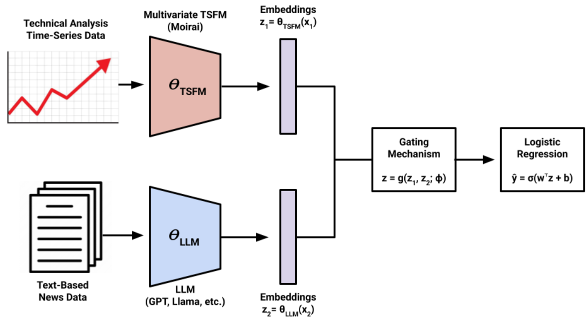

Figure 1: Multi-Model Model (M^3) for Market Sentiment Analysis

Traditional text analysis techniques often fail to capture the nuance and context inherent in natural language. Techniques such as dictionary-based or bag-of-words sentiment scoring fail to consider word order and syntactic structure, leading to oversimplification of financial text data (Mahajani et al., 2023).

Our work aims to cover several core areas of Natural Language Processing (NLP), primarily contextualized text representation, financial sentiment analysis, and multi-modal embedding integration. We focus on leveraging language models to derive rich, context-aware embeddings from financial text data, addressing the shortcomings of traditional methods like dictionary-based or bag-of-words approaches, which fail to capture linguistic nuance and syntactic structure.

We use transformer models to encode textual data from news articles and alerts, enabling more accurate sentiment inference and predictive modeling of stock returns. At the same time, we extend the scope of NLP to incorporate time-series modeling by integrating embeddings derived from financial technical analysis indicators, which are derived from stock HLOC prices and volume. For generating embeddings we use MoRAI, a masked autoencoder model for multi-modal forecasting (Woo et al., 2024). This reflects a cross-disciplinary bridge between NLP and time-series forecasting using transformer models.

One technical challenge we address is how to effectively integrate text and time-series data into a unified predictive framework. Our solution involves extracting embeddings separately from each modality and combining them via concatenation to enhance predictive performance. This enables the model to leverage both narrative market sentiment and quantitative technical signals, which are often treated in isolation in financial modeling tasks. By combining these modalities, we aim to improve the accuracy and robustness of a sentiment analysis baseline model to predict stock return and demonstrate the power of transformer models in a broader financial forecasting context.

2 Background

Our work builds on several key ideas and methods in the intersection of natural language processing and time-series forecasting. One foundational approach we draw from is the use of large language models (LLMs) for extracting contextualized representations from financial text data. For instance, Chen et al. (2022) demonstrate how LLMs can be leveraged to predict stock returns using representations of financial news. This highlights the growing value of textual analysis in financial prediction tasks. In our work we compare the BERT-base-uncased Masked Language Model and BAAI Sentence Transformer model for generating embeddings from the news data (Devlin et al., 2019) (Xiao et al., 2023).

Additionally, transformer-based architectures have shown increasing promise in time-series analysis, where their ability to model long-range dependencies makes them effective for forecasting. A notable example is Chronos-T5 (Ansari et al., 2024), a univariate time-series foundation model. For example in (Valeyre and Aboura, 2024), the Chronos-T5 model is used in its zero-shot and fine-tuned forms to forecast daily residual returns by predicting next-day returns. Furthermore, some multi-modal approaches for financial applications have been explored (Wang et al., 2023).

In our work, we combine these two paradigms: text-based embeddings from financial news and alert data, and time series-based embeddings from technical indicators. Our approach differs from the work of (Valeyre and Aboura, 2024) as we utilize a multi-modal approach leveraging both multivariate time-series and news data.

3 Data

Our multi-modal framework integrates two primary data sources to capture both market sentiment and technical trading signals: FPM Cloud and CRSP. We utilize the FMP Cloud News API to obtain financial news articles (Financial Modeling Prep, 2025). These articles are preprocessed and encoded using a transformer language model, which generates contextualized textual embeddings for each trading day t. This source was rated as best for real-time market data and news by (Singh, 2024). The news data from FMP Cloud fluctuates in quantity, especially after 2019. This is illustrated in Figure 2, and this influenced our decision to begin the out-of-sample evaluation of our model in 2019.

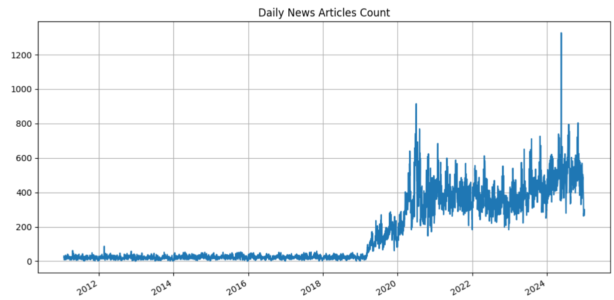

Figure 2: Daily News Articles from FMP Cloud

Historical stock price and volume data are obtained from the Center for Research in Security Prices (CRSP) using the Wharton Research Data Services (WRDS) (Wharton Research Data Services, 2025). Using the Pandas-TA open-source technical analysis library, we derive approximately 25 technical indicators from this time-series data. These indicators are input into Moirai which outputs latent embeddings that capture market dynamics up to day t.

For the top 1000 stocks by market equity per day, we extract ticker, date, adjusted price, HLOC (high, low, open, close) prices, cumulative 3-day returns, corresponding news data if available, and correpsonding news sentiment score for stocks between January 2011 and December 2024. The sentiment score label is generated using the following schema. We have a three-day window denoted by $t-1$, t, $t+1$. We generate the label for day t by the returns from day $t-1$ to $t+1$. If the returns are positive we assign a sentiment label of 1, otherwise we assign a sentiment label of 0. A table of the data is shown below:

DATE	MARKET_EQUITY	RET	TICKER	ADJ_PRC	close	open	high	low	3_DAY_RET	news_data	label
2017-12-29	2.845933	0.000000	DK	34.9400	34.94000	34.98000	35.22000	34.82000	0.021379	<Corresponding article data>	1
2017-12-29	2.840879	-0.023292	BXS	31.4500	31.45000	32.30000	32.30000	31.45000	-0.015577	<Corresponding article data>	0
2018-01-02	276947.656704	0.018252	GOOGL	21300.0000	1065.00000	1048.33997	1066.93994	1045.22998	0.033190	<Corresponding article data>	1
2018-01-02	229180.252615	0.016708	AMZN	23780.2002	1189.01001	1172.00000	1190.00000	1170.51001	0.015259	<Corresponding article data>	1
2018-01-02	193289.769696	0.030233	NVDA	7974.0004	199.35001	195.78000	199.50000	194.50000	0.076343	<Corresponding article data>	1

Figure 3: Snippet of Finalized Processed Data

4 Methods

We begin by training a baseline news-only model, which uses the BERT language mode to generate contextual embeddings from financial news headlines and articles. We then compare this model with a sentence-transformer known as BAAI, which was optimized for news sentiment analysis. These embeddings are passed to a logistic regression classifier that outputs a sentiment score for each stock, interpreted as the probability of positive short-term returns. These 2 baselines help to illustrate how the model performs without the time-series component, allowing us to show incremental value from our approach.

We then extend the architecture into a multimodal model that integrates both unstructured text and structured time-series data. For the textual component we use BAAI, and for the time-series component we use MoiRAI, we then concatenate our two output embeddings. This allows the model to weigh and combine information from both embedding sources more effectively, capturing complementary semantic signals. Finally, the combined embeddings are passed through a logistic regression model which we train to predict a sentiment score, representing the probability that this stock will yield positive returns in the next day. Notably, both of the embedding models remain frozen, and only the logistic regression layer is trained.

The out-of-sample simulated portfolio is constructed by taking long and short positions on the top 20% and bottom 20% of the stocks, ranked by sentiment scores. Each of the long and short lists are normalized so that they sum to 1 and -1, ensuring that the stronger sentiments get more weight, and that we have a net neutral portfolio. The algorithm is illustrated in Diagram 1. We have also included a model card explaining the details of our model, which is represented in figure 8 of the appendix.

Although we have technical indicator data for all stocks, embedding the entire universe daily is computationally infeasible given the time constraints. Therefore, we use the presence of news as a filtering step. Only stocks with available news are processed for both textual and technical embeddings. This dramatically reduces the number of stocks processed each day, making the approach scalable in practice.

We adopt a rolling window approach where the model is re-trained using data from the past 2 years (24 months) and then evaluated on the subsequent 3-month period. This process is repeated every 3 months, allowing the model to continuously adapt to changing market dynamics. We implement early stopping on the validation set during training, and shuffle the training data to ensure robust generalization of our model. A detailed list of training hyperparameters for the logistic regression model is shown in Table 1.

Table 1: Hyperparameters

Param	Value	Purpose
RETRAIN_FREQ	3	Retraining interval
MIN_SAMPLES	100	Minimum data per retrain
WINDOW_MONTHS	18	Rolling data window length
TRAIN_MONTHS	12	Training period
VAL_MONTHS	6	Validation period
BATCH_FRAC	0.05	Fraction of # train samples
BATCH_MIN/MAX	32 / 512	Batch size bounds
LR	5e-5	Learning rate
WD	1e-4	Weight decay
EPOCHS	50	Training epochs
PATIENCE	5	Early-stop patience
TOP_PCT	0.10	Top selection ratio
DROP_RATE	0.2	Dropout rate
EMBED_DIM	384	Embedding size
MOIRAI_DIM	768	Feature vector size
DROP_FUSE	0.3	Fusion-layer dropout
START_DATE	2019-01-01	Backtest start date

We develop a backtesting framework that iterates through each day available in our dataset, and simulate the performance of each strategy's portfolio. This process simulates how the model would be updated and deployed in a real world setting, allowing us to evaluate our strategy over historical data. Backtesting throughout the 5-year evaluation horizon allows us to obtain a sequence of out-of-sample predictions that can be used to evaluate our models, calculating various quantitative metrics for evaluation.

5 Results

Our multi-modal forecasting framework was evaluated over a 5-year period (2019–2024) using a monthly rolling window backtesting strategy. We assess all models on cumulative return, Sharpe ratio, maximum drawdown, and annualized volatility.

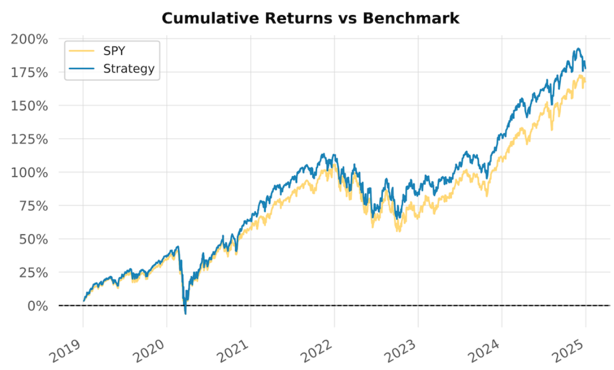

Figure 4: Equal-Weight Portfolio top 300 stocks by market equity (Baseline)

We begin with two baselines, visualized in Figure 3. The SPY index, acting as a passive benchmark, achieved a cumulative return of 167.77%, a Sharpe ratio of 0.93, and maximum drawdown of -33.72% with volatility of 19.8%. As expected, SPY delivered consistent, market-aligned performance.

The Equal-Weight (EW) Portfolio, which invests equally across the top 300 stocks by market equity, slightly outperformed SPY with a return of 177.70% and Sharpe of 0.96, though at a marginally higher drawdown of -35.13% and volatility of 19.96%.

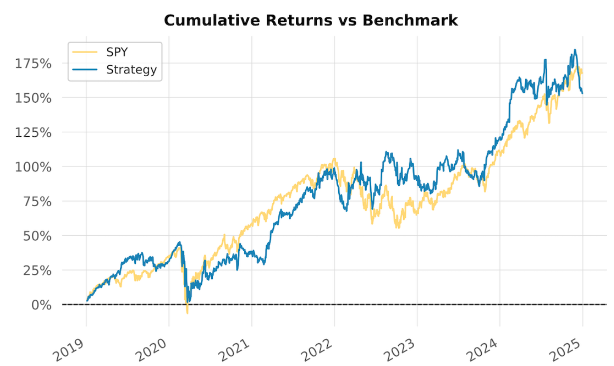

Figure 5: Text-based embedding model (BERT) to score sentiment (Baseline)

This presents our first text-based strategy, the BERT sentiment model. This model returned 152.99%, with a Sharpe of 0.83, max drawdown of -30.06%, and annualized volatility of 20.55%. BERT's structure gave it the most conservative risk profile among our models, but it did not outperform in cumulative returns or other important metrics. BERT struggled with ambiguity in complex headlines (e.g., "EPS beats expectations, but growth slows"), leading to inaccurate model predictions.

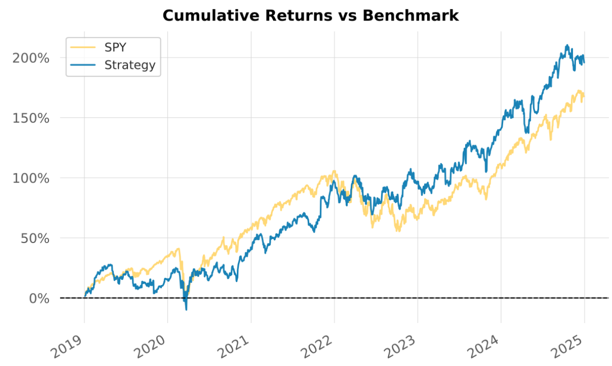

Figure 6: Text-based embedding model (BAAI Sentence Transformers) to score sentiment (Improved)

Figure 6 illustrates the BAAI Sentence Transformer model, which offered improved semantic comprehension. The model achieved 195.82% cumulative return, Sharpe of 0.95, and the lowest drawdown of -29.53% among all models, with volatility of 21.49%. This model demonstrates a strong improvement over bert, likely because it was designed for news sentiment analysis.

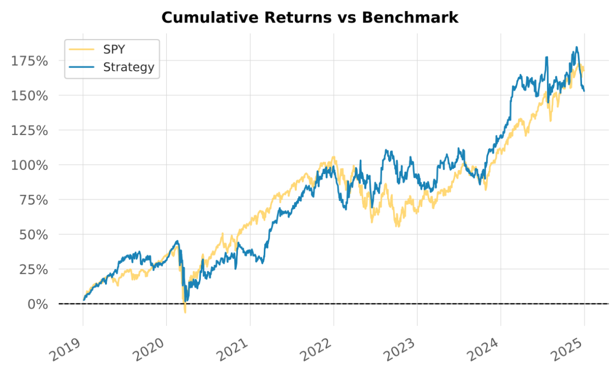

Figure 7: Multi-Modal model (M^3) for sentiment analysis using BAAI Sentence Transformers + Moirai TSFM (Improved)

The performance of our multi-modal model (M^3), is visualized in Figure 7. This model combines BAAI text embeddings with Moirai time-series embeddings. M^3 achieved a return of 173.21%, Sharpe ratio of 0.86, drawdown of -37.25%, and volatility of 22.36%, indicating a reduction in overall performance compared to the BAAI embeddings alone. Interestingly, the M^3 model is notably more volatile, likely due to the addition of the technical indicators.

The overall quantitative metrics used to evaluate our models is shown in table 2. This table summarizes the performance comaprison across all models and benchmarks. It is clear that the BAAI embeddings generate some alpha, demonstrating

Table 2: Performance Summary

Model	Ret. (%)	Sharpe	DD (%)	Vol. (%)
SPY	167.77	0.93	−33.72	19.82
EW	177.70	0.96	−35.13	19.96
BERT	152.99	0.83	−30.06	20.55
BAAI	195.82	0.95	−29.53	21.49
M^3	173.21	0.86	−37.25	22.36

that we can create profitable trading signals from news embeddings. It is clear that further experimentation is need to improve the M^3 approach, but the results show potential for improvement, perhaps with stronger embedding models or more complex gating mechanisms.

6 Conclusion

We introduced a novel multi-modal forecasting framework that integrates contextualized news embeddings with multivariate time-series embeddings to predict stock returns. By leveraging a transformer-based language models to encode daily news sentiment and the Moirai TSFM to capture historical time-series patterns, our approach demonstrates the capacity to generate actionable trading signals from real-time market data. Backtesting over a five-year period (2019–2024) revealed that the BAAI sentence transformer alone achieved the highest cumulative return (195.82%) and Sharpe ratio (0.95), outperforming traditional benchmarks such as the SPY Index and an equal-weight portfolio. Although our multi-modal model (M^3) exhibited greater volatility and modestly lower returns, it validated the feasibility of combining unstructured and structured data modalities within a unified predictive framework.

Future work will focus on enhancing the gating mechanism to more effectively combine news and technical embeddings, exploring alternative fusion strategies such as attention-based or nonlinear aggregation methods. Additionally, exploring the use of larger language models may yield better performance, as scaling laws can drastically increase text embeddings. Collectively, these extensions hold the potential to refine our multi-modal approach, and could advance the application of transformer models in quantitative finance.

References

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. 2024. Chronos: Learning the language of time series.

Yifei Chen, Brian Kelly, and Dacheng Xiu. 2022. Expected returns and large language models. Available at SSRN: https://ssrn.com/abstract=4416687.

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

Financial Modeling Prep. 2025. Stock News API Documentation. https://site.financialmodelingprep.com/developer/docs/stock-news-api. Accessed: 2025-05-10.

Siddhant Jaydeep Mahajani, Shashank Srivastava, and Alan F. Smeaton. 2023. A comparison of lexicon-based and ml-based sentiment analysis: Are there outlier words?

Pawan Singh. 2024. Best global news apis in 2025. Accessed: 2025-05-11.

Sebastien Valeyre and Sofiane Aboura. 2024. Llms for time series: an application for single stocks and statistical arbitrage.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention is all you need.

Ziao Wang, Yuhang Li, Junda Wu, Jaehyeon Soon, and Xiaofeng Zhang. 2023. Finvis-gpt: A multimodal large language model for financial chart analysis.

Wharton Research Data Services. 2025. Wharton Research Data Services (WRDS). https://wrds-www.wharton.upenn.edu/. Accessed: 2025-05-10.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. 2024. Unified training of universal time series forecasting transformers.

Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. 2023. C-pack: Packaged resources to advance general chinese embedding.

A Appendix

Model Card – M^3 Text2Cash Model

Model Details

- Developed to explore the integration of financial sentiment analysis and multimodal time series forecasting using transformer models.
- The model architecture involves:
  - **BAAI Model Embeddings** for encoding financial news articles
  - **MoRAI**, a masked autoencoder model, used to generate embeddings from technical indicators
  - A **logistic regression classifier** to predict stock sentiment

Intended Use

- Intended to analyze financial news and technical analysis data for predicting stock returns.
- Enables sentiment classification using both narrative market data and quantitative technical indicators.

Metrics

- Cumulative Returns, Sharpe Ration, Max Drawdown, Annualized Volatility

Training Data

- Text data from a corpus of financial news.
- Technical indicators extracted from public historical HLOC and volume data for selected stocks.

Evaluation Data

- We use a rolling window approach where the model is retrained using data from the past 2 years and then evaluated on the following 3-month period. This process is repeated every 3 months, advancing the window by one month each time.

Ethical Considerations

- Financial texts may contain biases; outputs may inherit or amplify these.
- No proprietary or sensitive data is used; sources are publicly available.

Caveats and Recommendations

- Model predictions are probabilistic and must not be treated as definitive investment advice.

Figure 8: Model Card