Leveraging Llama-3 Sentiment and FRED-Based Macroeconomic Indicators for Predictive FX Analytics

Md Raquib Hassan

Department of Electrical and Computer Engineering, North South University

Nikunja-2, Dhaka, Bangladesh[1]

raquib.hassan@northsouth.edu

Abstract— This paper presents an integrated framework for foreign exchange (FX) market prediction that combines sentiment signals from the Llama-3 language model with key macroeconomic indicators sourced from the Federal Reserve Economic Data (FRED). Unlike conventional machine learning methods that rely solely on numerical features, our approach incorporates contextual information from financial news and economic reports to enhance predictive power. The system constructs a sentiment index representing market optimism or pessimism toward specific currency pairs and fuses it with indicators such as inflation, interest rates, employment data, and GDP trends. We compare the performance of the Llama-3-based sentiment model with baseline machine learning techniques, including Random Forest, Support Vector Machines, and Gradient Boosted Trees. Results show that integrating sentiment features with FRED indicators yields improved forecasting accuracy, particularly in high-volatility environments. This framework demonstrates how large language models can complement traditional econometric and machine learning approaches to produce more robust and explainable FX analytics.

Keywords **— Foreign exchange prediction, FinBERT, LLaMA-3, sentiment analysis, financial news analytics, economic calendar events, macroeconomic indicators, FRED, multi-timeframe price data, machine learning, temporal forecasting, NLP-based financial modeling, FX markets.**

I. Introduction

Foreign exchange (FX) markets are influenced by macroeconomic fundamentals, investor sentiment, and geopolitical conditions. Price fluctuations are not solely driven by quantitative indicators such as inflation rates, GDP, or interest rate spreads; instead, market participants respond strongly to financial news, analyst commentary, and central bank statements. Traders interpret these textual sources emotionally, creating rapid momentum shifts. Traditional forecasting systems that rely only on historical price series or technical indicators are unable to capture this hidden sentiment dimension.

Recent developments in machine learning (ML) and natural language processing (NLP) provide tools for transforming financial text into quantifiable sentiment. Modern large language models (LLMs) such as Llama-3 allow contextual interpretation of financial content, which general-purpose sentiment models often misinterpret. An example is financial jargon like “hawkish policy,” “rating downgrade,” or “dovish stance,” which carries implied directional meaning. Bidirectional text modeling combined with trend mapping allows improved classification of whether the FX market will enter an uptrend or downtrend.*Text Font of Entire Document*

II. Literature Review

Text mining and NLP have become essential for extracting insights from financial texts. Yang et al. (2025) demonstrated this using a big-data–based model for financial institutions that applies NLP to classify and evaluate text-driven information. Their approach showed how unstructured content can support decision-making and value assessment; however, it focused on institutional text and did not address domain-specific market behaviors such as currency pair dynamics or macroeconomic dependencies.

Recent research has shifted toward domain-adapted large language models. Ballinari and Maly (2025) fine-tuned Llama-3.1 (8B) using FX news articles, addressing challenges such as specialized FX jargon and the relative nature of currency pairs. They proposed a hybrid labeling strategy combining human annotations with distant labels based on price movement and introduced separate “past” vs. “future” sentiment labels. Their work improved interpretability but lacked integration with fundamental indicators like interest rates or employment.

Large foundation models further advance financial analysis. FinTral, built on Mistral-7B, supports multimodal data and introduces FinSet, a benchmark covering nine financial tasks. Its training pipeline—domain pretraining, instruction tuning, and alignment—achieved strong zero-shot performance. However, its closed-source nature limits reproducibility and adoption in academic research.

The underlying Mistral-7B architecture (Jiang et al.) provides efficient long-context processing via grouped-query and sliding-window attention, outperforming larger open models while remaining accessible under Apache-2.0 licensing. These characteristics make it suitable for domain adaptation in financial NLP.

Temporal and relational approaches such as MI-FinText integrate sentiment, event detection, and value prediction using multi-task learning and temporal graph networks. While powerful for corporate or equity analytics, it does not address FX-specific characteristics such as monetary policy, macro indicators, and directional trend movement.

Overall, literature shows three core gaps: (1) limited FX-specific sentiment models, (2) weak integration of unstructured sentiment with macroeconomic fundamentals, and (3) insufficient temporal modeling for trend prediction. This work addresses these gaps by combining Llama-3 sentiment analysis with FRED-based economic indicators to classify FX trends as uptrend or downtrend.

III. METHODOLOGY

This study introduces a multi-modal forecasting framework designed to predict short- and medium-term foreign exchange (FX) market reactions following macroeconomic announcements. The framework integrates three key data modalities: structured economic-calendar information, multi-timeframe FX price series, and sentiment extracted from FX news headlines using FinBERT and LLaMA-3. The methodological pipeline consists of data acquisition, preprocessing and alignment, sentiment extraction, multi-timeframe feature construction, and predictive modeling.

A. Data Acquisition

Macroeconomic announcement data were collected from publicly available economic-calendar providers. Each entry contains the event timestamp (UTC), event name, country, category, importance level, and numerical fields including actual, forecast, and previous values. This dataset serves as the primary source of macroeconomic drivers.

Historical EUR/USD price data were exported from the MetaTrader 5 (MT5) platform using three time resolutions. Five-minute (M5) candles capture microstructure responses around announcements, one-hour (H1) candles provide the main predictive horizon, and daily (D1) candles reflect broader macroeconomic trends. Each candle includes open, high, low, close (OHLC), tick volume, spread, and real volume.

A complementary dataset of FX-related news headlines was collected from financial news sources. Each headline includes a timestamp and serves as an unstructured information source to model market psychology. These titles enable extraction of sentiment and contextual meaning, which conventional numerical indicators cannot capture.

B. Data Preprocessing

All timestamps were standardized to UTC. Because MT5 uses broker server time, candle timestamps were converted accordingly, and news headlines were aligned to a ±3-hour window around each economic event to ensure meaningful temporal association. Missing values in the economic calendar, particularly forecast or previous values, were imputed using the most recent known data. Binary indicators were added to preserve the information content associated with missingness.

For each macroeconomic event, forward returns were computed to quantify FX market reaction. Let (Pt) denote the EUR/USD price nearest to the event release; the forward return over horizon (Δ t) is defined as:

$$R_{\Delta t} = \frac{P_{t+\Delta t} - P_t}{P_t}.$$

Reaction labels were computed for four horizons: 5 minutes, 1 hour, 3 hours, and 24 hours after each event.

C. Sentiment Extraction from FX News Titles

Two complementary NLP models were used to extract sentiment and semantic meaning from news headlines. FinBERT, a BERT-based transformer fine-tuned specifically on financial text, was used to classify each headline into positive, negative, or neutral categories, and to generate sentiment probability scores. This model captures finance-specific terminology and tone more accurately than general-purpose sentiment models.

To complement FinBERT's polarity scores, LLaMA-3 was used to generate high-dimensional semantic embeddings from each headline. These embeddings capture deeper contextual cues such as macroeconomic tone, risk-on or risk-off sentiment, and forward-looking expectations. Embeddings associated with each event were aggregated to create a sentiment-context vector aligned with the event. News titles appearing within the defined temporal window of each event were mapped to that event, ensuring that both FinBERT-based sentiment scores and LLaMA-3 embeddings contributed to the predictive feature set.

D. Multi-Timeframe Feature Engineering

To construct a unified dataset, features from the 5m, 1H, and 1D price series were synchronized with event timestamps. The M5 data provide short-term features such as micro-returns, volatility, liquidity proxies, and momentum prior to each announcement. The H1 data deliver medium-term signals including trend strength, ATR-based volatility, candlestick patterns, and rolling statistical measures. The D1 data supply long-term context such as macro trend indicators, volatility regimes, and extended momentum.

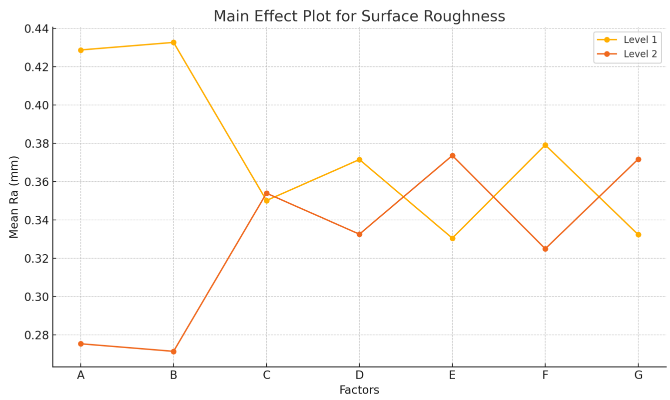

Fig. 1 A sample graph

Additional features derived from the economic calendar include numerical surprise measures (actual minus forecast), percentage surprise, event category, importance level, country encoding, and missing-value indicators. Sentiment features consist of FinBERT polarity probabilities, LLaMA-3 semantic embeddings, and the temporal density of news coverage surrounding each event. All engineered features are merged into a single event-aligned dataset.

E. Predictive Modeling

Two forecasting models were developed. XGBoost was employed as a baseline due to its effectiveness on tabular data and its ability to provide interpretable feature importance metrics.

TABLE I

Font Size	**Appearance (in Time New Roman or Times)**		
	Regular	**Bold**	**Italic**
8	table caption (in Small Caps), figure caption, reference item		reference item (partial)
9	author email address (in Courier), cell in a table	abstract body	abstract heading (also in Bold)
10	level-1 heading (in Small Caps), paragraph		level-2 heading, level-3 heading, author affiliation
11	author name		
24	title		

The primary model, a Temporal Fusion Transformer (TFT), was selected for its ability to integrate heterogeneous feature types—including multi-resolution time series, categorical variables, and sentiment embeddings—while capturing temporal dependencies and enabling multi-horizon forecasts. The combined feature set, consisting of price-based, sentiment-based, and fundamental indicators, was used as input to the TFT.

F. Evaluation

The dataset was chronologically split into training, validation, and testing subsets to avoid look-ahead bias. Performance evaluation included metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), directional accuracy, and category-specific accuracy (e.g., CPI, NFP, GDP). Ablation tests were conducted to assess the contribution of FinBERT sentiment features and LLaMA-3 embeddings. Model interpretability was examined through TFT attention weights and XGBoost feature importance.

The version of this template is V2. Most of the formatting instructions in this document have been compiled by Causal Productions from the IEEE LaTeX style files. Causal Productions offers both A4 templates and US Letter templates for LaTeX and Microsoft Word. The LaTeX templates depend on the official IEEEtran.cls and IEEEtran.bst files, whereas the Microsoft Word templates are self-contained. Causal Productions has used its best efforts to ensure that the templates have the same appearance.

Fig. 2 shows an example of a low-resolution image which would not be acceptable, whereas Fig. 3 shows an example of an image with adequate resolution. Check that the resolution is adequate to reveal the important detail in the figure.

Please check all figures in your paper both on screen and on a black-and-white hardcopy. When you check your paper on a black-and-white hardcopy, please ensure that:

- the colors used in each figure contrast well,
- the image used in each figure is clear,
- all text labels in each figure are legible.

Fig. 3 shows an example of a low-resolution image which would not be acceptable, whereas Fig. 3 shows an example of an image with adequate resolution. Check that the resolution is adequate to reveal the important detail in the figure.

Please check all figures in your paper both on screen and on a black-and-white hardcopy. When you check your paper on a black-and-white hardcopy, please ensure that:

- the colors used in each figure contrast well,
- the image used in each figure is clear,
- all text labels in each figure are legible.
- Paper both on screen and on a black-and-white hardcopy.

Please check all figures in your paper both on screen and on a black-and-white hardcopy.

APPENDIX

The preferred spellingo word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression, “One of us (R. B. G.) thanks ” Instead, try “R. B. G. thanks”. Put sponsor acknowledgments in the unnumbered footnote on the first page. References are important to the reader; therefore, each citation must be complete and correct. If at all possible, references should be commonly available publications

ACKNOWLEDGMENT

The preferred the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression, “One of us (R. B. G.) thanks . . .” Instead, try “R. B. G. thanks”. Put sponsor acknowledgments in the unnumbered footnote on the first page. References are important to the reader; therefore, each citation must be complete and correct. If at all possible, references should be commonly available publications

REFERENCES

[1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.