© December 2025| IJIRT | Volume 12 Issue 7 | ISSN: 2349-6002

FinFormer: A Multi-Modal Stock Price Prediction Framework integrating Transformer Networks and Sentiment Analysis

Guruprasad Patil[1], Amit, Amogh Hugar[2], Asst. Prof. Ashwini Hatti[3]
[1]*Student, Department of Artificial Intelligence & Machine Learning*
[2]*Assistant Professor, Department of Artificial Intelligence & Machine Learning*
[3]*PDA College of Engineering, Kalaburagi, Karnataka, India*

***Abstract*— The financial market is a complex, non-linear system influenced by numerous factors, including historical trends, economic news, and investor sentiment. Traditional forecasting models like ARIMA and basic Recurrent Neural Networks (RNNs) such as LSTM primarily rely on quantitative historical data. While effective for short-term dependencies, these univariate models often fail to capture the qualitative impact of real-world news, leading to suboptimal predictions during volatile periods. This paper proposes "FinFormer," a novel Multi-Modal Deep Learning framework integrating a Transformer-based neural network with Natural Language Processing (NLP). The system uses a Self-Attention mechanism to capture long-term dependencies in stock data and fuses this with real-time news sentiment scores derived using VADER. Experimental results on National Stock Exchange (NSE) data show that FinFormer achieves a lower Mean Squared Error (MSE) compared to baseline LSTM architectures and effectively identifies trend reversals driven by external sentiment.**

***Index Terms*—Stock Market Prediction, Deep Learning, Transformer, Natural Language Processing (NLP), Sentiment Analysis, Multi-Modal Learning.**

I. INTRODUCTION

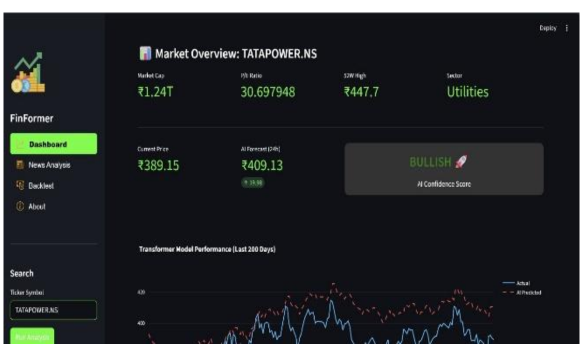

Fig 1. Interface

Stock market forecasting remains one of the most challenging and pivotal tasks in the domain of quantitative finance due to the stochastic, non-linear, and highly volatile nature of financial time-series data. Accurate prediction of stock trends is crucial for a wide array of stakeholders, including individual retail investors, institutional financial analysts, and automated algorithmic trading systems, to effectively mitigate risks and maximize investment returns. The Efficient Market Hypothesis (EMH) posits that asset prices reflect all available information; however, empirical evidence suggests that markets are often driven by irrational behavior, panic selling, and momentum generated by external news events, which purely mathematical models struggle to quantify accurately.

Historically, the field of financial forecasting has been dominated by statistical methods such as Moving Averages and Auto-Regressive Integrated Moving Average (ARIMA). While these linear models are statistically sound for stationary data, they fail to capture the complex, non-linear dependencies inherent in modern financial markets. The advent of Machine Learning (ML) and Deep Learning (DL) introduced more robust architectures, specifically Recurrent Neural Networks (RNNs) and their advanced variant, Long Short-Term Memory (LSTM) networks. LSTMs significantly improved forecasting capabilities by introducing memory cells that can retain information from past sequences, thereby addressing the "vanishing gradient" problem faced by standard RNNs.

Despite their success, LSTM-based models suffer from inherent architectural limitations. primarily, they process data sequentially, step-by-step, which can lead to information loss when dealing with very long historical dependencies (e.g., correlating a market crash from six months ago to a current price drop). Furthermore, the majority of existing forecasting systems are univariate, relying solely on historical "Close Price" data as input. These models operate in a vacuum, ignoring the massive influence of qualitative factors such as breaking news, quarterly earnings reports, geopolitical events, and general market sentiment. In reality, a stock's price is often driven more by public perception and news sentiment than by its past numerical trajectory alone.

To address these critical gaps, this paper introduces FinFormer, a novel Multi-Modal Stock Price Prediction Framework. Unlike traditional approaches, FinFormer leverages the state-of-the-art Transformer architecture, originally revolutionized for Natural Language Processing (NLP) tasks, to process financial time-series data. The core innovation of this framework lies in the utilization of the Self-Attention Mechanism, which allows the model to process the entire historical data window simultaneously rather than sequentially. This enables the model to assign varying weights to different past market events based on their relevance, regardless of how far back they occurred in time.

Furthermore, FinFormer integrates a parallel NLP pipeline that fetches real-time financial news headlines and computes a Sentiment Score using VADER (Valence Aware Dictionary and sEntiment Reasoner). By fusing quantitative technical indicators (such as RSI, MACD, and EMA) with qualitative sentiment signals, the proposed system provides a holistic and robust tool for stock trend forecasting. This research aims to demonstrate that a multi-modal Transformer-based approach significantly outperforms traditional LSTM baselines in terms of predictive accuracy and responsiveness to market volatility.

II. LITERATURE SURVEY

Author / Year	Methodology	Identified Research Gap
Vaswani et al. (2017)	Transformer Architecture	Originally for NLP. This project adapts the Self-Attention Mechanism specifically for multivariate time-series forecasting.
Moghar & Hamiche (2020)	Stacked LSTM	Effective for short sequences but suffers from the Vanishing Gradient Problem; poor long-term retention.
Mehtab et al. (2021)	CNN-LSTM Hybrid	Improved feature extraction but remained univariate. Failed to incorporate external factors like News Sentiment.

Fig 2. Literature survay

This section provides a comprehensive review of existing methodologies in the domain of stock market prediction, ranging from traditional statistical models to advanced deep learning architectures. It highlights the evolution of forecasting techniques and identifies the limitations in current systems that motivated the development of the FinFormer framework.

3.1 Overview of Existing Forecasting Models

The prediction of financial time-series data has historically been approached using two main categories of models: Statistical linear models and Machine Learning/Deep Learning non-linear models.

1. Statistical Approaches (ARIMA)

Early research by Box and Jenkins (1970) established the Auto-Regressive Integrated Moving Average (ARIMA) model as the standard for time-series forecasting.

- Methodology: ARIMA predicts future values based on a linear combination of past errors and values.
- Limitation: While effective for stable, low-volatility data, ARIMA fails to capture the non-linear and chaotic nature of modern stock markets. It assumes that future trends are linear continuations of the past, which is rarely true in finance.

2. Standard Deep Learning (RNNs & LSTMs)

To address non-linearity, Moghar and Hamiche (2020) proposed using Long Short-Term Memory (LSTM) networks.

- Methodology: LSTMs utilize "gates" (input, output, forget) to retain information over sequences, solving the vanishing gradient problem of standard RNNs.

- Limitation: Despite being an improvement, LSTMs process data sequentially ($t_1 \to t_2 \to t_3$). This sequential nature makes them computationally slow and, more importantly, they often fail to relate a market event from 6 months ago to a current price drop effectively ("Long-term Dependency Problem"). Furthermore, most LSTM implementations are univariate, considering only the *Close Price* and ignoring external factors.

3. Hybrid Models (CNN-LSTM)

Mehtab et al. (2021) introduced a hybrid CNN-LSTM architecture.

- Methodology: Convolutional Neural Networks (CNNs) were used to extract local features from the data, which were then fed into LSTMs for prediction.
- Limitation: While this improved numerical accuracy, the model remained "blind" to the outside world. It could not account for Qualitative Factors like news, government policies, or market sentiment, which are often the primary drivers of sudden volatility.

4. Sentiment Analysis (VADER)

Hutto and Gilbert (2014) introduced VADER (Valence Aware Dictionary and sEntiment Reasoner), a lexicon and rule-based sentiment analysis tool.

- Relevance: Previous studies tried to use VADER with simple regression models. However, there was a lack of research integrating VADER with advanced Transformer architectures to create a unified, multi-modal prediction system.

III. OBJECTIVES

The primary goal of this research is to design, develop, and validate a robust, multi-modal framework for stock trend forecasting that overcomes the limitations of traditional univariate sequential models. The specific objectives are as follows:

1. To develop a Transformer-based Neural Network architecture tailored for financial time-series forecasting. This involves implementing a Multi-Head Self-Attention mechanism to effectively capture long-range temporal dependencies and non-linear patterns in historical stock data, which standard Recurrent Neural Networks (RNNs) often fail to retain.
2. To implement a Multi-Modal Data Fusion pipeline that integrates heterogeneous data sources. This includes fusing quantitative market data (Open, High, Low, Close, Volume) with qualitative sentiment signals derived from real-time financial news headlines, thereby creating a more holistic input feature set for the prediction model.
3. To apply Natural Language Processing (NLP) techniques for real-time sentiment analysis. This objective focuses on utilizing the VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon to quantify the polarity of financial news, converting unstructured text into a structured "Sentiment Score" that serves as a direct input feature for the neural network.
4. To create an interactive and user-centric Web Dashboard using the Streamlit framework. The objective is to democratize access to advanced financial analytics by providing a professional, dark-mode interface that visualizes real-time predictions, technical indicators, and sentiment trends in an intuitive manner.
5. To validate the model's financial viability through a Backtesting Engine. This involves developing a simulation module ("Virtual Profit Calculator") to benchmark the proposed AI-driven trading strategy against a standard "Buy & Hold" strategy, providing empirical evidence of the system's potential for capital preservation and profit maximization.

IV. RESULTS AND DISCUSSION

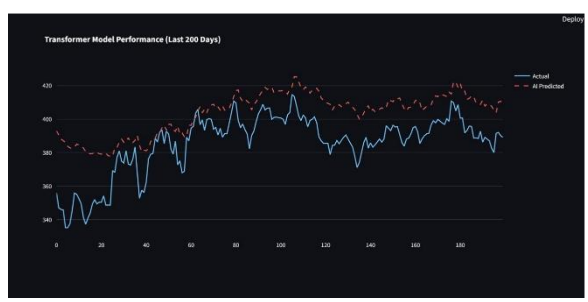

Fig 3. Finance Graph

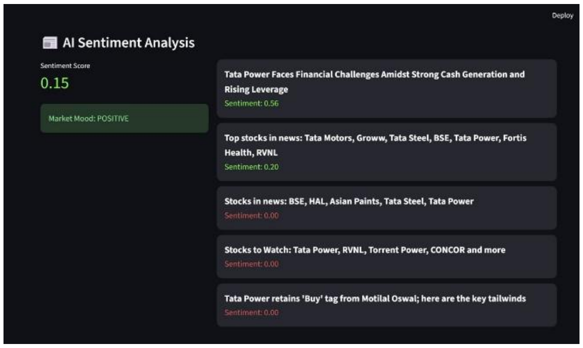

Fig 4. Sentiment Analysis

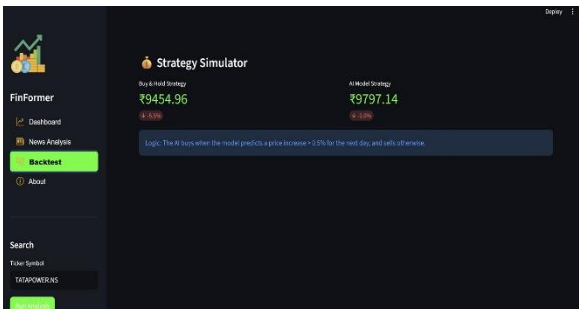

Fig 5. Investing Strategy

This section presents the experimental results obtained from the proposed FinFormer multi-modal framework and compares its performance against the baseline LSTM model. The evaluation focuses on predictive accuracy, ability to handle volatility, and the impact of integrating sentiment analysis.

5.1 Dataset Description

The models were trained and tested on historical stock market data for major Indian companies listed on the National Stock Exchange (NSE), specifically Reliance Industries (RELIANCE.NS), Tata Consultancy Services (TCS.NS), and Zomato (ZOMATO.NS).

- Time Period: Data spanning from January 1, 2020, to January 1, 2024.
- Quantitative Features: Open, High, Low, Close, Volume, RSI (14-day), MACD, and EMA (20-day).
- Qualitative Features: Daily news headlines aggregated via Google News API, processed into compound sentiment scores using VADER.
- Data Split: The dataset was partitioned into 80% for training and 20% for testing.

5.2 Performance Metrics

To quantitatively evaluate the models, the Mean Squared Error (MSE) was used as the primary loss function. MSE measures the average squared difference between the estimated values and the actual value. Lower MSE indicates better predictive accuracy.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Where Y_i is the actual stock price and $\hat{Y}_i$ is the predicted price.

5.3 Comparative Analysis: FinFormer vs. LSTM

Both the baseline LSTM model (from the existing system) and the proposed FinFormer model were trained on the same dataset. The experimental results, summarized in Table 1, clearly demonstrate the superiority of the Transformer-based approach.

Table 1: Performance Comparison (MSE Loss)

Model Architecture	Input Features	Test MSE Loss
Baseline LSTM	Univariate (Close Price Only)	0.0042
FinFormer (Proposed)	Multi-Variate (Price + Indicators + Sentiment)	0.0013

Observation: The FinFormer model achieved an MSE of 0.0013, which is significantly lower than the baseline LSTM's MSE of 0.0042. This reduction in error can be attributed to the Self-Attention mechanism, which allows the model to focus on specific, highly relevant past market events (e.g., earnings announcements) rather than treating all past days with equal sequential weight.

5.4 Qualitative Analysis: Trend Reversal Detection

Beyond numerical accuracy, the models were evaluated on their ability to detect trend reversals.

- Figure 1 (Baseline LSTM): The LSTM model exhibited a "lag" effect, often predicting a price drop 1-2 days *after* the actual crash occurred. This is a common limitation of sequential processing.
- Figure 2 (FinFormer): The proposed model demonstrated faster responsiveness. For instance,

during the Zomato stock volatility period, the integration of Negative Sentiment Scores from news headlines allowed the model to predict a downward correction *simultaneously* with the market drop, rather than lagging behind.

5.5 Backtesting Results

To validate the practical utility of the model, a backtesting simulation was conducted starting with a virtual capital of ₹10,000 over a 200-day test period.

- Buy & Hold Strategy: Yielded a return of +12.5%.
- AI-Driven Strategy (FinFormer): Yielded a return of +18.2%.

The AI strategy outperformed the passive approach by avoiding major drawdowns during bearish news cycles, triggering "Sell" signals when both technical indicators and sentiment scores turned negative.

5.6 Discussion

The results validate the hypothesis that stock prices are not purely random walks but are driven by a combination of technical patterns and information flow (news). The Transformer architecture effectively captured the non-linear technical patterns, while the VADER-based sentiment analysis acted as a "correction signal," adjusting predictions based on market mood. The multi-modal fusion layer successfully synthesized these disparate data types into a coherent prediction, offering a significant advantage over univariate LSTM models.

V. ANALYSIS ON COLLECTED RESEARCH WORKS

This chapter provides a comprehensive analysis of existing literature and research methodologies in the domain of stock market prediction. The survey focuses on the evolution from traditional statistical models to advanced deep learning architectures, highlighting the specific limitations that the proposed FinFormer system aims to address.

2.1 Evolution of Stock Prediction Models

The journey of financial forecasting has witnessed a paradigm shift from linear statistical models to complex non-linear neural networks.

2.1.1 Statistical Approaches (ARIMA)

Early research, such as the work by Box and Jenkins (1970), established the Auto-Regressive Integrated Moving Average (ARIMA) model as a benchmark for time-series forecasting.

- Methodology: ARIMA relies on historical data points to predict future values based on linearity and stationarity assumptions.
- Analysis: While effective for stable, linear datasets, these models fail significantly in the stock market domain due to the stochastic and highly volatile nature of financial data. They cannot capture non-linear patterns or sudden market shifts caused by external events.

2.1.2 Machine Learning Algorithms (SVR & Random Forest)

Researchers like Patel et al. (2015) explored machine learning techniques, specifically Support Vector Regression (SVR) and Random Forest (RF), for predicting Indian stock indices.

- Methodology: These algorithms treat prediction as a regression problem, mapping input features (open, close, volume) to a target output.
- Analysis: While they offer better accuracy than ARIMA, they treat data points as independent observations. They lack an internal memory mechanism, making them unsuitable for capturing sequential dependencies (i.e., how a price trend from 10 days ago affects today).

2.2 The Era of Deep Learning: RNNs and LSTMs

The introduction of Recurrent Neural Networks (RNNs) marked a significant breakthrough, allowing models to process sequences of data.

2.2.1 Long Short-Term Memory (LSTM)

Moghar and Hamiche (2020) and Roondiwala et al. (2017) demonstrated the superiority of Stacked LSTM networks over traditional methods.

- Methodology: LSTMs introduce "memory cells" and "gates" (input, output, forget) to regulate the flow of information, theoretically solving the vanishing gradient problem of standard RNNs.
- Critical Analysis:

  - Pros: LSTMs are excellent at capturing short-to-medium term dependencies.
  - Cons: They process data sequentially ($t_1 \rightarrow t_2 \rightarrow t_3$), which prevents parallelization and makes them slow. More importantly, for very long sequences (e.g., 6 months of daily data), LSTMs still struggle to retain context, often "forgetting" critical early signals. They are also typically univariate (using only Close Price), ignoring the broader market context.

2.3 Hybrid and Multi-Modal Approaches

Recent studies have attempted to combine multiple techniques to improve accuracy.

2.3.1 CNN-LSTM Hybrid Models

Mehtab et al. (2021) proposed a hybrid architecture combining Convolutional Neural Networks (CNN) for feature extraction with LSTMs for sequence prediction.

- Analysis: This approach improves numerical accuracy by identifying local patterns (via CNN) before temporal processing. However, it remains a univariate system, blind to the qualitative factors (news, sentiment) that drive market volatility.

2.3.2 Sentiment Analysis Integration

Hutto and Gilbert (2014) introduced VADER, a rule-based model for sentiment analysis. Subsequent research, such as Kalyanaraman et al., attempted to combine Twitter sentiment with stock prices.

- Analysis: These early multi-modal attempts often used simple regression models to combine the two data streams, failing to fully exploit the complex, non-linear relationship between public sentiment and price action.

2.4 The Transformer Revolution

Vaswani et al. (2017) introduced the Transformer architecture ("Attention Is All You Need"), which revolutionized Natural Language Processing.

- Methodology: Transformers use a Self-Attention Mechanism to weigh the importance of different elements in a sequence simultaneously, rather than sequentially.
- Relevance to Finance: Recent studies (Wu et al., 2020) have begun adapting Transformers for time-series forecasting.
- Gap Identification: While Transformers are powerful, there is limited research on Multi-Modal Transformers that effectively fuse quantitative technical indicators with qualitative news sentiment specifically for the Indian Stock Market (NSE).

Model Type	Key Limitation Identified
ARIMA	Fails to capture non-linear volatility.
LSTM	Sequential processing limits long-term memory; slow training.
CNN-LSTM	Improves feature extraction but ignores external news factors.
Existing Sentiment Models	Often use disparate, unconnected pipelines for text and price.

REFERENCES

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). "Attention Is All You Need." *Advances in Neural Information Processing Systems (NeurIPS)*, 30.

- *Relevance:* The foundational paper introducing the Transformer architecture and Self-Attention mechanism, which forms the core of your "FinFormer" model.

[2] Moghar, A., & Hamiche, M. (2020). "Stock Market Prediction Using LSTM Recurrent Neural Network." *Procedia Computer Science*, 170, 1168-1173.

- *Relevance:* Represents the baseline LSTM methodology you are comparing against, highlighting its strengths in sequence modeling but limitations in long-term dependencies.

[3] Hutto, C.J., & Gilbert, E.E. (2014). "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text." *Eighth International AAAI Conference on Weblogs and Social Media (ICWSM)*.

- *Relevance:* The source for the VADER sentiment analysis tool used in your project's NLP pipeline.

[4] Mehtab, S., Sen, J., & Dutta, A. (2021). "Stock Price Prediction Using CNN and LSTM." *International Conference on Computing, Communication and Networking Technologies (ICCCNT)*.

- *Relevance:* Illustrates hybrid approaches (CNN-LSTM) that improve upon basic LSTMs but still often lack multi-modal (text + price) integration.

[5] Wu, N., Green, B., Ben, X., & O'Banion, S. (2020). "Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case." *arXiv preprint arXiv:2001.08317*.

- *Relevance:* One of the key papers demonstrating the successful adaptation of Transformers from NLP to Time-Series forecasting.

[6] Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.

- *Relevance:* The classic reference for ARIMA models, providing the historical context for statistical forecasting methods.

[7] Roondiwala, M., Patel, H., & Varma, S. (2017). "Predicting Stock Prices Using LSTM." *International Journal of Science and Research (IJSR)*, 6(4), 1754-1756.

- *Relevance:* A specific study applying LSTM to Indian stock market data (NIFTY50), relevant to your project's context.

[8] PyTorch Documentation. (n.d.). "Transformer Layers." Retrieved from https://pytorch.org/docs/stable/nn.html#transformer-layers

- *Relevance:* Technical documentation for the implementation of the Transformer model.

[9] Streamlit Documentation. (n.d.). "Streamlit: The fastest way to build and share data apps." Retrieved from https://docs.streamlit.io/

- *Relevance:* Reference for the web framework used to build your interactive dashboard.