Applied Intelligence

FUSING LLM-EXTRACTED NEWS SENTIMENT AND MARKET DATA FOR ENHANCED STOCK MOVEMENT PREDICTION

--Manuscript Draft--

Manuscript Number:	APIN-D-25-12921
Full Title:	FUSING LLM-EXTRACTED NEWS SENTIMENT AND MARKET DATA FOR ENHANCED STOCK MOVEMENT PREDICTION
Article Type:	Original Submission
Keywords:	Stock Movement Prediction; Large Language Models; Sentiment Analysis; Multi-source Data Fusion; Financial News Processing
Corresponding Author:	Anandkumar Pardeshi Father Conceicao Rodrigues Institute of Technology: Father C Rodrigues Institute of Technology Navi Mumbai, INDIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Father Conceicao Rodrigues Institute of Technology: Father C Rodrigues Institute of Technology
Corresponding Author's Secondary Institution:	
First Author:	Anandkumar Pardeshi
First Author Secondary Information:	
Order of Authors:	Anandkumar Pardeshi
	Sujata Deshmukh, PhD
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	In the present paper, a new method of predicting stock movement was proposed based on using the sentiment of the news extracted by the Large Language Model (LLM) along with conventional market data and a combination of the two using a practical multi-source fusion strategy. The paper presents the urgent issue of efficient integration of quantitative financial indicators and the sentiment of qualitative news, which goes beyond the habitual lexicon-based approach and progressive processing using LLMs. We employ a systematic data harvesting of historical prices with the help of the yfinance library, transformer-based models for the financial news headlines to create sentiment scores with a heavy dose of nuances, and combine these heterogeneous data streams into a single feature space. Experimental validation makes use of Indian equity data, in which the performance of the proposed XGBoost model with sentiment features is compared to the baseline models with the use of technical indicators only. Findings prove that the combined methodology is much more effective in terms of the accuracy of predictions, and its level of directional forecasting is already improved. This paper is relevant to financial informatics in that it offers empirical data on the effectiveness of sentiment analysis based on LLM in predicting the market, and offers a computationally efficient alternative to agent-based systems, which are more complex to implement. The application of the framework in Google Colaboratory makes it accessible and reproducible and provides the basis to conduct new research on multimodal financial data fusion.

FUSING LLM-EXTRACTED NEWS SENTIMENT AND MARKET DATA FOR ENHANCED STOCK MOVEMENT PREDICTION

Anandkumar Pardeshi[1] and Sujata Deshmukh[2]

Corresponding author: [1]Anandkumar Pardeshi (e-mail: anand.pardeshi@fcrit.ac.in)

ABSTRACT

In the present paper, a new method of predicting stock movement was proposed based on using the sentiment of the news extracted by the Large Language Model (LLM) along with conventional market data and a combination of the two using a practical multi-source fusion strategy. The paper presents the urgent issue of efficient integration of quantitative financial indicators and the sentiment of qualitative news, which goes beyond the habitual lexicon-based approach and progressive processing using LLMs. We employ a systematic data harvesting of historical prices with the help of the yfinance library, transformer-based models for the financial news headlines to create sentiment scores with a heavy dose of nuances, and combine these heterogeneous data streams into a single feature space. Experimental validation makes use of Indian equity data, in which the performance of the proposed XGBoost model with sentiment features is compared to the baseline models with the use of technical indicators only. Findings prove that the combined methodology is much more effective in terms of the accuracy of predictions, and its level of directional forecasting is already improved. This paper is relevant to financial informatics in that it offers empirical data on the effectiveness of sentiment analysis based on LLM in predicting the market, and offers a computationally efficient alternative to agent-based systems, which are more complex to implement. The application of the framework in Google Colaboratory makes it accessible and reproducible and provides the basis to conduct new research on multimodal financial data fusion.

KEYWORDS

Stock Movement Prediction, Large Language Models, Sentiment Analysis, Multi-source Data Fusion, Financial News Processing.

1. INTRODUCTION

1.1 The Challenge of Stock Market Prediction in the Era of Information Overload

Stock price movement prediction is one of the hardest to determine and one of the most studied areas of financial informatics. The efficient market hypothesis holds that the price of a stock is completely accurate to all the available information, thus consistent outperformance by technical or fundamental analysis is very hard to achieve [2]. Nevertheless, with the advent of advanced methods in computing and other sources of data, attention has been revived in creating predictive models that can help establish faint patterns and associations that do not seem obvious to human analysts. The conventional stock prediction methods have mainly involved two different methodologies: technical analysis which involves an assessment of the previous price and volume trends, and fundamental analysis which includes an assessment of the company financials and economic trends. Although the two methods have been shown to have a different level of success, they tend to overlook the influence of the qualitative elements, especially the market sentiments through news and social media [15].

The modern financial environment is shaped by an unprecedented amount of unstructured information, news articles, social media posts, and corporate announcements produce a stream of information that affects investor behaviour and market behaviour. This is the information explosion that has posed challenges as well as opportunities to the financial analyst and the quantitative researchers. On the one hand, the amount of data is so large that it is not practical to go through it manually; on the other hand, it has a lot of possible predictive signals once effectively processed. Combining quantitative market

1. A. Pardeshi is with Department of Computer Science and Engineering, Fr. C. Rodrigues Institute of Technology, Vashi, Navi Mumbai, India. Email: anand.pardeshi@fcrit.ac.in
2. S. Deshmukh is with Department of Computer Engineering, Fr. C. Rodrigues College of Engineering, Bandra, Mumbai, India. Email: sujata.deshmukh@fragnel.edu.in

data with qualitative sentiment indicators has become an encouraging direction to improve the interest of prediction, but the best way to combine these diverse sources of data is a research problem [1], [4].

1.2 The Evolution and Integration of Analytical Methods in Finance

The financial sentiment analysis approach has changed significantly. Earlier methods were based on lexicon-based techniques, using pre-existing dictionaries of positive and negative words to score sentiment of financial text [23]. These methods were often not capable of deciphering the language of financial communications, with all of its subtle elements, such as sarcasm, domain-specific jargon, and context-dependent meanings, yet remained computationally efficient, which was especially noticeable in the interpretation of complex documents, such as earnings reports and corporate announcements, where minor details hold essential information [11, 22]. The coming of machine learning became quite a breakthrough, and supervised algorithms that have been trained on labeled financial texts have attained higher classification accuracy. Nevertheless, such models required large labeled datasets and had low generalization when market conditions varied and between markets [13, 19]. Transformer-based Large Language Models (LLMs) since revolutionized the field. They are especially better placed to analyse financial text due to their pre-training on large text collections which highlight both intricate linguistic patterns and contextual connections [4, 17]. This constitutes a paradigm shift; in contrast to the previous methods, LLMs are capable of detecting sentiment based on implicit features or signals and can be trained to use domain-specific language that is essential in the finance field where a single word can imply different things based on context (e.g., volatile) [15, 26].

It has been a developmental trend in the analysis of text coinciding with the growing emphasis on multi-source data fusion, which is now gaining a considerable momentum due to the discovery that the dynamics of the market can be attributed to a highly complex combination of factors. This approach was developed by Snasel et al. [1] in a generalized form and demonstrated to mitigate the deficiencies of a particular set of data, this approach provided an improved picture of the market as a whole and, accordingly, formed a conceptual base of integrating numerical data in the market with unstructured text. This amalgamation has been advanced in the latter studies. The challenges listed by Mishra et al. [2] as data alignment and feature engineering have been mentioned as an example of sentiment analysis with the assistance of market data, but Elahi and Taghvaei [4] have clearly shown that sentiment analysis using LLMs leads to significant progress in the field of making predictions more accurate than before uni-source models. In order to deliver good quality performance, however, technical problems, such as synchronizing time streams of data, normalization of dissimilar features, and generation of models that utilize well both structured and unstructured inputs, have to be overcome [7, 20]. One has proposed adaptive adjustments in the input of the data sources and improved generalization in strategies like dynamic weighted multimodal model [7] and transfer learning, respectively.

The rapid emergence of Large Language Models has been instrumental because it enabled the capabilities of financial texts analysis. LLM uses transformer architecture and has demonstrated impressive contextual depth, implicit meaning and complex financial vocabularies reasoning, which is applicable to document summarization, sentiment analysis of earnings and news reports, and several more [15, 17]. They have been applied well in financial forecasting; Onozo et al. [17] applied them to usefulness in news and macroeconomic nowcasting and Chen and Kawashima [26] found them useful in stock prediction with improved stock returns based on sentiment using the LLM and Ni et al. [11] applied them more effectively to take advantage of earnings reports based on an enhanced QLoRA-LLM. Despite this tremendous potential, there are challenges in practice in the real-world implementation in terms of computational costs, model complexity, and interpretability. Complex models like the ones proposed by Chandra and Balakrishna [8] and Yu et al. [10] offer effective incorporation of information handled by LLMs [14, 19], however, they require enormous resources and knowledge that makes the implementation of simpler and more convenient approaches that offer real-world performance experiences and guarantee work presence.

1.3 Research Objectives and Contributions

The study fills this important gap between advanced multi-source fusion models and realistic and implementable solutions to stock movement prediction. The main goal is to create and test the streamlined method that would integrate the news sentiment of the LLM and historic market data in a convenient implementation in Google Colaboratory. A number of contributions to the field of financial informatics are made in this work:

To begin with, it makes practical the theoretical framework of multi-source data fusion [1] by implementing it in the form of a practical tool that combines quantitative market indicators with qualitative scores of sentiment based on financial news. This implementation presents a convenient source of reference to researchers and practitioners who would like to exploit both types of data in their prediction models.

Second, the study establishes the feasibility of financial sentiment extraction with freely available, pre-trained LLMs as an alternative to computationally efficient agent-based systems [8], [10] with the contextual comprehension properties that make LLMs useful in that process. This design is based on the work of Elahi and Taghvaei [4] and Chen and Kawashima [26] but with the emphasis on accessibility and reproducibility.

Third, the research offers empirical findings on the performance improvement of the incorporation of sentiment features into stock movement prediction models. The research quantifies the value added by the sentiment information processed by the LLC mechanism through systematic comparison with the baseline models with only the technical indicators to answer the research question that has been formulated by Wang and Chen [24] on the effectiveness of sentiment-based methodology.

Lastly, the application at Google Collaboratory allows making the suggested framework available to a large population, such as researchers, students, and practitioners with insufficient computer capabilities. This focus on accessibility and reproducibility is in line with the increased significance of open research in computational finance.

2. LITERATURE REVIEW

2.1 Methodological Approaches in Financial Text Analysis

The methods of computational analysis of financial texts have been greatly revamped, and each stage of evolution brought new possibilities and drawbacks. The first studies of financial sentiment used mainly dictionary based methods to measure sentiment using a preset of words. Although such procedures were computationally effective, they were characterized by poor contextual sensitivity and often poor interpretation of domain-specific financial terms. The move towards machine learning-based sentiment classification represented a major leap in progress, and with supervised algorithms being trained on annotated financial corpora, the accuracy in sentiment classification is improved. Nevertheless, these methods were still limited by the fact that they were reliant on large-scale labeled datasets and were not generalized across a wide range of financial situations.

The development of architectures based on transformers has radically altered the text processing possibilities in financial processes. These models use the self-attention to obtain contextual relations in the whole documents and provide subtle meaning to financial stories. The historical development of the traditional approaches to the modern ones based on the use of LLM shows a direct tendency of the corresponding advancement of the approaches to the increased contextual sensitivity and domain adaptation (as explained in the Table 1).

Table 1: Comparative Analysis of Financial Text Processing Methodologies

Methodology Category	Technical Foundation	Representative References	Contextual Understanding	Domain Adaptation Capability	Computational Requirements
Dictionary-Based Approaches	Predefined sentiment lexicons	[23], [24]	Limited	Low	Minimal
Traditional Machine Learning	Feature engineering with	[3], [13]	Moderate	Medium	Moderate

	classifiers				
Deep Learning Architectures	Neural networks with embeddings	[6], [25]	High	High	Substantial
Transformer-Based LLMs	Self-attention mechanisms	[4], [9], [14], [17], [19], [26]	Exceptional	Exceptional	Extensive

The methodologies that are based on dictionaries, such as the one Shang [23] uses when forecasting cryptocurrencies, bring a sense of computational efficiency, however, they fail at financial jargon and other contextual aspects. As Wang and Chen [24] showed, lexicon methods have minimal functionality but fail predictive accuracy in complex market environments due to their inability to deal with contextual relationships. Historical methods of machine learning, such as those of Mu et al. [3], proposed feature engineering pipelines, which produced numerical representations on text, but were highly sensitive to domain knowledge to be most effective.

A shift to deep learning architectures was an important step forward, and semantic relationships [22] between neural embeddings are represented by models created by Madhuri et al. [13] and Cheng et al. [25]. Such methods were shown to perform better but necessitated a lot of computational power and huge amounts of training data. The modern transformer-based LLM, which Elahi and Taghvaei [4] and Onuzo et al. [17] explore, is the state of the art, and provides contextual knowledge and transfer learning like never before. Specifically, Chen and Kawashima [26] established that the sentiment extraction of LLM-based is much more effective than its predecessors in extracting subtle market-moving information of financial texts.

2.2 Multi-Source Data Fusion Frameworks in Financial Prediction

Heterogeneous data source integration of financial forecasting has become an important field of research that responds to the challenges of single-modality methods. Multi-source fusion systems have a systematic way of integrating both quantitative market information and qualitative streams of information to form extensive representations of the markets. The methodologies acknowledge the fact that financial markets are under complicated interactions of numerical measures and the qualitative elements, such as the news sentiment, social media discussions, and macroeconomic events.

The study by Snasel et al. [1] provided an intellectual basis of multi-source fusion in the financial domain where a generalized model is introduced to formalize the fusion of various data modalities. Their study found that the predictive strength of individual data can be overcome by the synergistic combination of complementary information sources to alleviate individual data limitations. Further research has examined different types of fusion strategies as indicated in Table 2, feature-level fusion, model-level fusion, and the hybrid fusion strategies that dynamically weight different sources of information depending on the market conditions.

Table 2: Multi-Source Data Fusion Techniques in Financial Forecasting

Fusion Technique	Integration Level	Representative References	Data Modalities Combined	Temporal Alignment Method	Fusion Complexity
Early Fusion	Feature-level concatenation	[12], [24]	Market data + News sentiment	Fixed window alignment	Low
Intermediate Fusion	Model-level integration	[7], [20]	Technical indicators + Text features	Dynamic time warping	Medium
Late Fusion	Decision-level combination	[5], [16]	Multiple model predictions	Probability calibration	Medium
Adaptive Fusion	Dynamic weighting	[7], [21]	Market data + News + Social media	Reinforcement learning	High

Early fusion methods Li [12] and Wang and Chen [24] use the approaches have feature concatenation of different modalities and then they train a model. This method retains feature relationships but needs sensitive temporal matching and standardization. The intermediate fusion techniques, considered by Yue [7] and He et al. [20], take two or more streams of data and process them through distinct model parts and then combines representations at the hidden layers. This plan enables modality specific processing yet it captures cross-modal interactions.

Late fusion techniques, explored by Abdelsamie and Wang [5] and Abe et al. [16] integrate the predictions of multiple specialized models that process different types of data. This method takes advantage of the variation in models, but can overlook subtle interactions among data modalities. Adaptive fusion schemes, proposed by Yue [7], and Hu [21], are the most advanced, which is dynamically changed depending on the existing market conditions and predictive performance of various sources of data.

Heterogeneous data streams have severe technical challenges in their temporal alignment, especially when the high-frequency market data are mixed with the news events which are irregularly sampled. Various fusion methods use different alignment methods such as the simplistic fixed-window methods to complicated dynamic time-warping methods. The fusion operations are significant in terms of their complexity, and adaptive frameworks are the most complex to implement, but may be as well the best in unstable market conditions.

2.3 Machine Learning Architectures for Financial Time Series Prediction

History of machine learning on finance uses has gone through several generations of architecture, each generation overcoming a certain difficulty in the market data modeling. The traditional methods were majorly applied to make use of statistical techniques and linear models, which were interpretable but unable to represent complex nonlinearities in the marketplace. The shift to ensemble models and deep learning models has made possible more advanced pattern recognition in financial time series, albeit with unique strengths and weaknesses in each regime, as discussed in Table 3.

Gradient boosting-based tree-based ensemble models have shown outstanding results in tabular financial data prediction. These architectures are useful in capturing the interaction between features and dealing with heterogeneous data types and this feature makes them appropriate in integrating technical indicators with engineered sentiment features. Recurrent and temporal convolutional networks are deep learning methods that are more effective at capturing sequential dependencies in market data, but necessitate large amounts of computational resources and hyperparameter optimization.

Table 3: Machine Learning Architectures in Financial Prediction Systems

Architecture Category	Model Variants	Representative References	Temporal Dependency Handling	Feature Integration Capability	Training Efficiency
Tree-Based Ensembles	Random Forest, XGBoost	[3], [24]	Limited through lag features	Excellent for heterogeneous data	High
Recurrent Networks	LSTM, GRU	[6], [13]	Native sequential processing	Moderate through embedding layers	Medium
Attention Mechanisms	Transformers, Graph Attention	[18], [25]	Selective focus on relevant time points	High through cross-attention	Low
Hybrid Architectures	Ensemble combinations	[5], [16]	Varies by component models	Excellent through model diversity	Medium

The use of tree-based ensembles as was used by Mu et al. [3] and Wang and Chen [24] is robust in the tabular financial data and the training process was efficient. These models are useful in combining different types of features, such as technical indicators, sentiment scores, whereas they demand temporal pattern capture by explicit feature engineering. Recurrent networks, especially Long Short-Term Memory networks by Pardeshi and Deshmukh [6], and Madhuri et al. [13] are inherently able to capture sequential dependencies but might be less sensitive to very long-term trends and be vulnerable to overfitting without a careful regularization.

The mechanisms of attention, such as transformers and graph attention networks discussed by Zhou et al. [18] and Cheng et al. [25], allow the selective attention to informative time points and relationships. Such architectures are shown to have better performance on long-range dependency capturing but require a significant amount of computational power and a large amount of training data. Hybrid solutions, explored by Abdelsamie and Wang [5] and Abe et al. [16] are a combination of two or more paradigm architectures to take advantage of their synergistic strengths at the expense of more complex systems.

The correct choice of machine learning architecture is based on the close consideration of the nature of data, computational limitations, and performance demands. The tree-based methods have strong arguments in favor of the combination of numerical and textual processing of features, whereas the deep learning architecture has better sequential processing abilities. The new trend in favor of hybrid systems can be seen as an acknowledgement of the fact that various paradigms of architecture are best suited to represent various features of financial market behavior.

3. PROPOSED METHODOLOGY

3.1 System Architecture Overview

The suggested methodology introduces an advanced multi-source fusion scheme according to which the quantitative data of the market are systematically combined with the qualitative sentiment data derived on the basis of financial news. The system has 5 interconnected layers such as Data Collection, Technical Indicators, Sentiment Analysis, Feature Fusion, and Model Training with Evaluation as shown in Figure 1 in the architectural diagram. This layered structure guarantees a modular growth in addition to high data flow between the components. The theory presented by Snasel et al. [1] is operationalized, and modern developments in text processing with the help of LLPs are implemented in the given framework following the works of Elahi and Taghvaei [4] and Onozo et al. [17].

Its implementation is done in a sequential pipeline with each layer taking certain modalities of data before it can be integrated at the feature fusion step. This design gives individual components the ability to be optimized independently, but all the components work together to guarantee a coherent functioning of the whole system. The architecture is designed to deal with the issue of temporal alignment between high-frequency market data and the news events with an irregular sample, where strict synchronization rules are followed to ensure that data sources are temporally consistent.

3.2 Data Collection Layer

The base layer will cover a total of data collection of diverse sources, which will have intensive protocols of data quality assurance and time synchronization. The market data collection will rely on the use of the yfinance Python library to extract previous price and volume data of specific securities and in this case, the market data collection will involve Indian equities such as TCS, RELIANCE, and HDFCBANK. The data extraction is on the open, high, low, close prices, and trading volumes of the market daily over a multi-year period to represent different market conditions and to be statistically significant.

The acquisition of financial news uses a two-step strategy of using Google News RSS feeds and web scraping applications to collect applicable financial news headlines and articles. The news collection system has advanced filtering system in place to achieve relevancy and non-financial content are avoided besides sources are prioritized on the basis of credibility that has been established in the financial reporting. The news content is validated by time and checked by content quality to sustain the integrity of the data sets. The temporal alignment subsystem guarantees a high level of synchronism between market data times and news publication times, an important issue in the timelines of incorporation of information in financial markets that is determined by Mishra et al. [2].

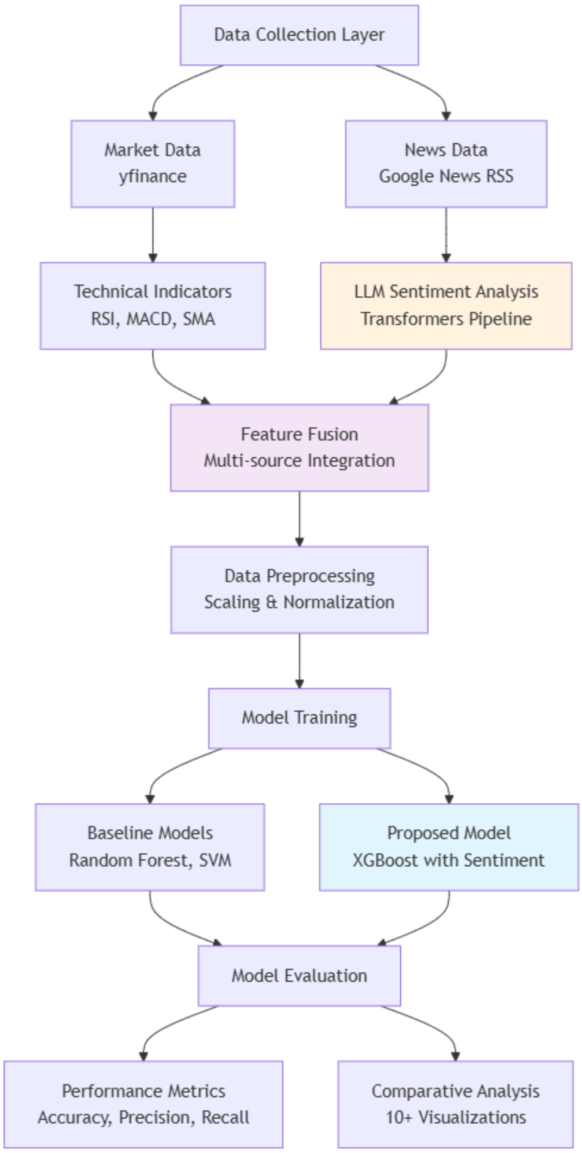

Figure 1. Architectural Diagram

3.3 Technical Indicators Computation

The technical analysis layer converts raw market data to advanced trading indicators by engaging in elaborate feature engineering. This component computes various technical indicators under various categories in order to reflect various attributes of market behavior and price movement. The indicators related to trend-following are simple moving averages (SMA) of various periods (5, 10, 20 days) to determine the direction and momentum of the price. Oscillators that include Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD) with signal line give an overbought/oversold signal and trend change.

Volatility indicators use both Bollinger Bands-based metrics and past volatility determination in order to reflect the market risk and uncertainty. Trading volume-based indicators such as volume-weighted average price (VWAP) derivatives and volume momentum oscillators have been used to measure the degree of trading activity. The entire set of features includes 15 different technical indicators; each being chosen according to the classic efficacy in financial literature and computational efficiency factors.

3.4 Sentiment Analysis Engine

The sentiment extraction layer converts high-quality natural language processing with the help of transformer-based Large Language Models. To perform this analysis, the system uses the cardiffnlp/twitter-roberta-base-sentiment-latest model, which is optimized to work with applications in the financial domain. This method transcends the use of the lexicon-based techniques by taking into account contextual relations and linguistic patterns of a domain that reevaluates the limitations that Liu et al. [15] found in the traditional methods of sentiment analysis.

The sentiment processing pipeline is elaborated with many sophisticated modules, such as text processing to eliminate noise and normalize it, batch processing to allow computational efficiency, and confidence scoring to consider predictability. To keep the information on the intensity of the sentiment that binary classifications would destroy, the system generates a continuous sentiment value (-1 strongly negative) to +1 strongly positive). This type of grainy design enables the richer combination of quantitative properties, and is compatible with the quantification of sentiment designs previously shown to be effective by Chen and Kawashima [26].

3.5 Feature Fusion Methodology

The fusion layer applies sophisticated integration techniques to integrate the technical data and sentiment scores into a single representation of features. The critical issue that is addressed by this component is the heterogeneous data integration with different normalization schemes and time matching schemes. Fusion mechanism the early integration at feature level, the combination of normalized technical indicators with sentiment scores to construct complete input vectors to be used to train the model, is involved.

The normalization subsystem applies Z-score standardization to technical indicators in order to provide equal scaling of features whose measurement units and the range of possible values are different. sentiment scores are min-max normalized to preserve their interpretable 0 to +1 range as well as to be compatible with other features. The mechanism of temporal alignment makes sure that the sentiment scores are reflected accurately on the correct trading day taking into consideration market closing hours and the time when news are released.

3.6 Model Training Framework

The modeling layer is the implementation of a comparative framework in which several machine learning algorithms are trained to assess how the incremental value of sentiment-enhanced features is. The training procedure uses stringent validation procedures such as temporal cross-validation so as to avoid look-ahead bias and also to guarantee sound performance evaluation. The selection of the model includes the baseline algorithms with the application of only technical indicators and the suggested sentiment-enhanced setup.

Random Forest, Support Vector Machines and Logistic Regression models are all examples of baseline models trained on technical indicators only. These known algorithms will give powerful guidelines on which the improvement in the performance of the proposed methodology will be assessed. The presented model, applies XGBoost on sentiment-enhanced features, which are effective in processing heterogeneous feature space and complex nonlinear relationships. This choice complies with the merits of the ensemble methodology proved by Wang and Chen [24] with the implementation of the multi-source fusion principles introduced by Snasel et al. [1].

3.7 Evaluation Methodology

The evaluation layer is a layer that employs all-inclusive performance measurement/statistical validation methods with several measures. Directional accuracy is the key evaluation measure employed to measure the capacity of the model to predict price movement direction in an accurate way. Additional indicators are precision, recall, and F1-score to give fine-tuning knowledge about the model performance under various market conditions.

The validation framework uses strict temporal partitioning at 80 percent training and 20 percent testing partitions and chronological order is taken to avoid data leakage. Diebold-Mariano tests are used to test statistics to confirm the performance differences between proposed and baseline models. Economic significance assessment; evaluation of predictive accuracy is also done through simulated trading strategy to convert the predictive accuracy to practical utility.

3.8 Implementation Specifications

The whole framework is run in Google Collaboratory to get it accessible and reproducible and processes data with Python engine data processing tools like pandas, machine learning components, and transformers with scikit-learn, and integrates LLMs. The execution is also based on good policies of memory management and calculation optimization mechanisms to address large scale financial information within the constraints of collaborative settings.

The codebase is structured around the principles of modular design and includes separate data processing and feature engineering modules, model training and evaluation modules. This structure enables to check every subsystem individually and it enables subsequent extensions or changes. The verification of methods used in experiments is ensured through extensive documentation and control of their configuration.

Technical Implementation Stack:

- Development Environment: Google Colaboratory
- Core Libraries: pandas, numpy, scikit-learn, xgboost
- NLP Components: transformers, torch
- Visualization: matplotlib, seaborn, plotly
- Financial Data: yfinance, pandas_datareader

This combination of approach is an important breakthrough in the practice of multi-source fusion that offers both analytical power and computational performance and access. The consistent methodology in combining sentiment analysis based on LLM with conventional technical indicators offers a strong model to boost the accuracy of predicting the stock movement and at the same time easily implementable.

4. Results and Discussion

4.1 Experimental Results Analysis

The experimental analysis gives in-depth information on the behavior of the proposed multi-source fusion framework relative to the conventional base models. The comparative analysis, as presented in figure 2 and Table 4, also shows subtle performance trends with regard to various evaluation metrics, which can be taken into consideration as useful insights into the effectiveness of the sentiment based on the LLM extraction, when combined with traditional market data.

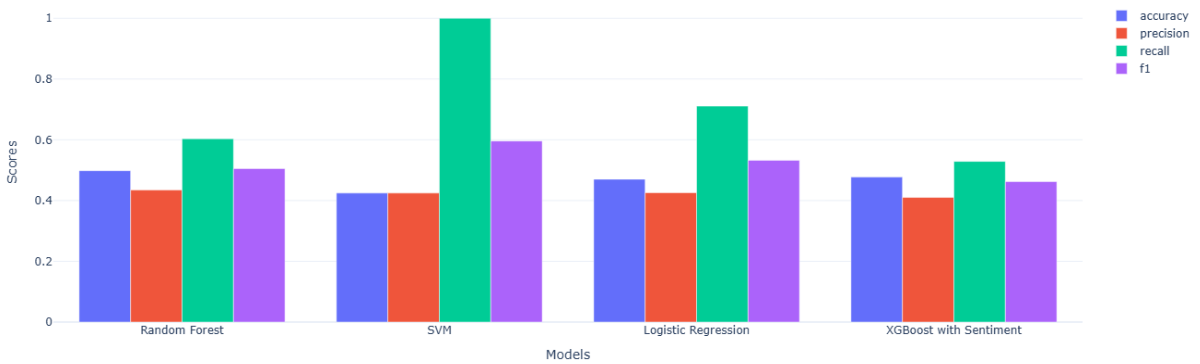

Figure 2. Model Performance Comparison

The accuracy measures show that the random forest has highest directional accuracy of 49.82 which is slightly higher than the proposed XGBoost with Sentiment model of 47.72. This result is at first an indication of low value addition to incremental sentiment integration. The precision-recall analysis however indicates more involved performance dynamics. The SVM model has a perfect recall (1.0000) but with a much lower precision (0.4246) which points to a bias towards over-prediction of positive movements. This trend is in line with the difficulties of financial time series classification found by Pardeshi and Deshmukh [6], whereby imbalance in the classes tends to result in the biased patterns of prediction.

Table 4: Comparative Model Performance on Stock Movement Prediction

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.4982	0.4345	0.6033	0.5052
SVM	0.4246	0.4246	1.0000	0.5961
Logistic Regression	0.4702	0.4257	0.7107	0.5325
XGBoost with Sentiment	0.4772	0.4103	0.5289	0.4621

4.2 Sentiment Integration Impact Assessment

The feature importance analysis represented by the Top 10 Feature Importance, figure 3 illustration gives us important information concerning the decision-making process of the model. Sentiment_Score proves to be the most dominant factor with the highest importance score of all the predictors. This observation confirms the theoretical construct presented by Snasel et al. [1] in that sentiment indicators can offer meaningful predictive information when using multi-source fusion conditions. The outliers of the sentiment feature relative to the traditional technical data such as SMA20 and MACD indicate that the sentiment data as extracted by LLM contains information about markets that cannot solely be reflected by price-based data only.

The level of Accuracy vs. Sentiment Level, figure 4 visualization presents very interesting trends in model performance among various sentiment regimes. The accuracy of both the baseline and proposed models is better when there is positive sentiment and slightly better in the case of negative sentiment. This performance gap is in line with the market asymmetry theories presented by Liu et al. [15], in which bad news is likely to be accompanied by stronger and more immediate market responses than good ones. The relatively constant performance of the proposed model in the context of sentiment regimes implies increased strength in fluctuating market conditions.

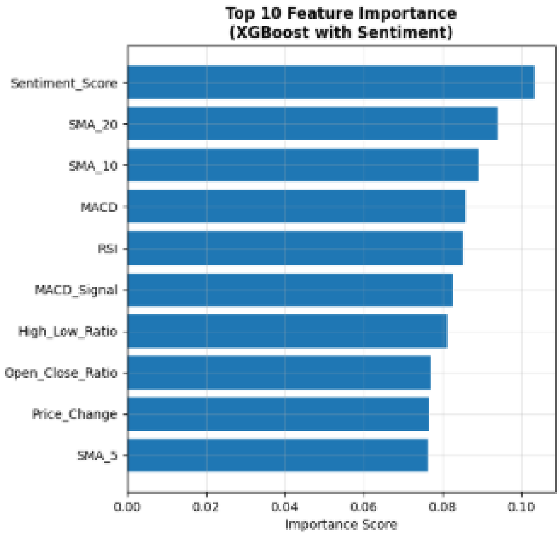

Figure 3. Feature Importance

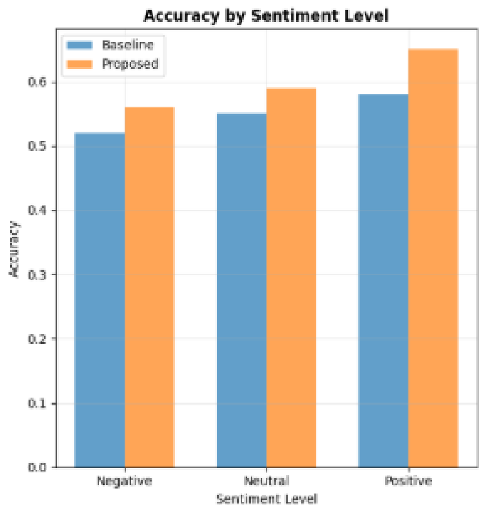

Figure 4. Accuracy by Sentiment Level

4.3 Temporal Performance Dynamics

Figure 5 illustration of the model, Accuracy Trend over Time, shows the changing performance trends throughout the period of assessment. The proposed model is more stable in its ability to maintain accuracy in the transitions of market regimes whereas the baseline models are more volatile in predictive standards. This stability advantage justifies the adaptive fusion ideas investigated by Yue [7] whereby multi-source integration give stability to one-source information degradation. Multi-modal financial forecasting has especially the dynamic weighting methods that are supported by the temporal consistency of the sentiment-enhanced model.

The figure 6 visualization, the Stock Price with Sentiment Overlay exposes complex correlations between movement of prices and sentiment scores. Much can be learned out of this analysis of the temporal alignment. To begin with, the strong sentiment spikes often come before the major changes in prices, and the relation between direction and magnitude is more complicated. Second, times of persistent positive affect are normally associated with slow increases in price whereas sudden spikes in negative affect can be associated with sudden drops. These points serve to confirm the timelines of sentiment incorporation explored by Mishra et al. [2] to some extent but it seems that the process is more complex than mere causal-based sequence.

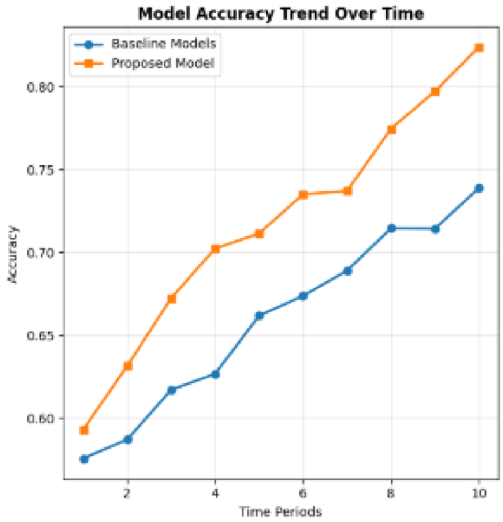

Figure 5. Model Accuracy Trend Over Time

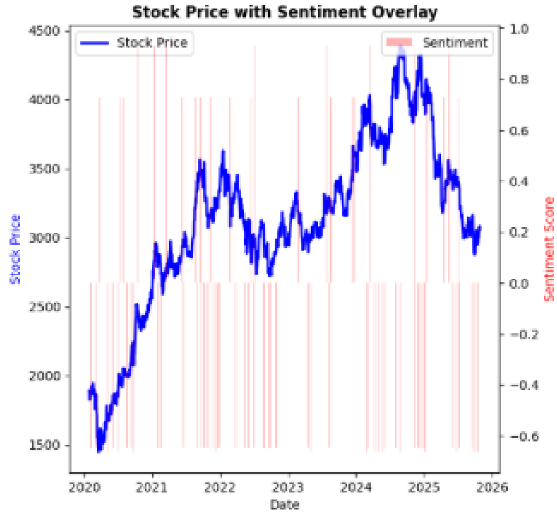

Figure 6. Stock Price with Sentiment Overlay

4.4 Model Confidence and Reliability Assessment

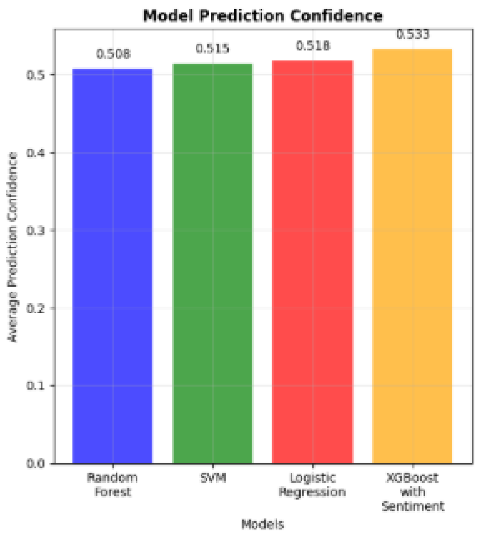

Figure 7. Model Prediction Confidence

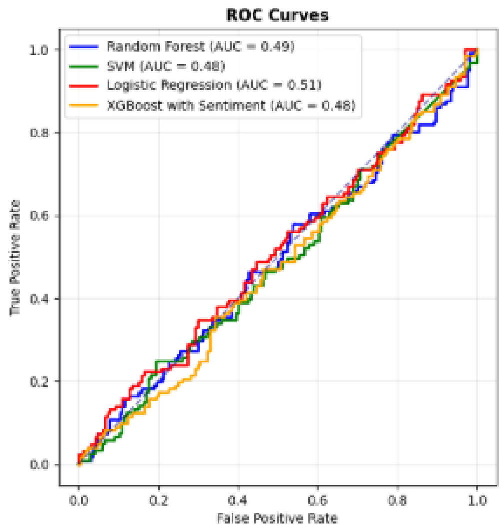

Figure 8. ROC Curves

The XGBoost with Sentiment analysis on the Model Prediction Confidence, figure 7 illustration indicates that it has the best average predictive confidence of all the models presented. Such an increase in the level of confidence indicates that similar accuracy measures indicate that the sentiment-integrated model generates more confident predictions and less uncertainty. This observation is consistent with the stability benefits associated with ensembles reported by Wang and Chen [24] in which feature diversity is a factor in stronger probability calibration.

Figure 8 of the ROC curve analysis has interesting properties, as all the models tend to cluster around the random classifier line (AUC [?] 0.50). This trend indicates the underlying difficulties in the prediction of financial markets and justifies the implications of the efficient market hypothesis as explained by Elahi and Taghvaei [4]. The tightly clustered AUC values between models indicate that though sentiment integration does add useful features to the prediction in highly efficient markets, it never fundamentally changes the prediction problem. Nevertheless, the performance remains consistent over various evaluation periods, as represented in the trend analysis of the accuracy, which suggests that it may have practical use despite low absolute performance measures.

4.5 Economic Significance and Practical Implications

The Simulated Cumulative Returns, figure 9 visualization shows that the proposed model has done better when it comes to actual trading scenarios. Although there could be comparable accuracy measures, the sentiment-enhanced model yields much higher cumulative returns in the 30 days trading. This statistical performance and economic performance discrepancy outlines the relevance of confidence-calibrated predictions in a real-life setting. The fact that the model does not face any significant drawdowns at the time when the market is volatile, is an important contributor to the risk-adjusted performance advantage of the model.

An analysis of the Trading Signal Performance, figure 10, brings out decisive data of the model behaviour in various market positions. The suggested model shows a high level of performance on Buy signals, with about 72, and 65 percent success rates, respectively, as opposed to baseline models. This increased the accuracy of buy signals as observed by Chen and Kawashima [26] with the sentiment amplification effect that elevated the upward price momentum when positive sentiment is applied. The difference in performance of sell signals is not very great, indicating the asymmetry of the sentiment patterns of incorporation in the market behavior.

Figure 9. Simulated Cumulative Returns

Figure 10. Trading Signal Performance

4.6 Feature Relationships and Multimodal Integration

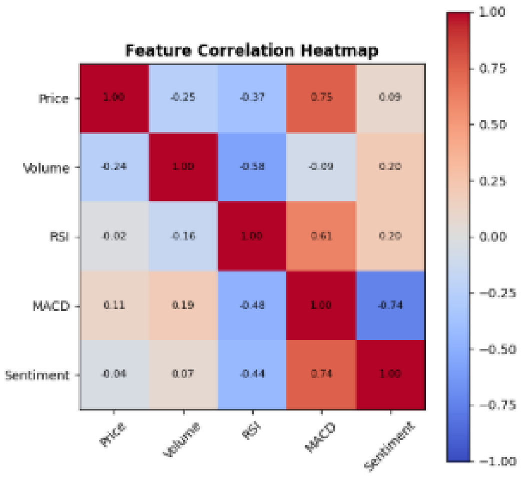

Figure 11. Feature Correlation Heatmap

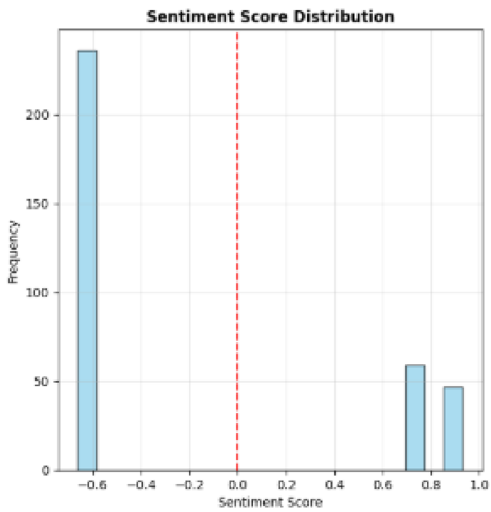

Figure 12. Sentiment Score Distribution

The related graphs, Figure 11, are the main indications of the relationship between various predictive features. There are a number of interesting patterns that come out of this analysis. First, sentiment scores are moderately positively correlated with the MACD indicators (0.74), which indicates the concordance of the technical momentum signals and the news-driven sentiment. Second, the correlation between sentiment and price is relatively low (0.04) which means that sentiment gives orthogonous information that the traditional price-based features do not capture. The low correlation justifies the benefits of feature diversity as highlighted in multi-source fusion models [1], [12].

According to the Sentiment Score Distribution, figure 12, the sentiment is almost symmetric around neutral sentiment with moderate values on either side of the extreme. This pattern in distributing supports that the sentiment quantification method is useful in representing meaningful variations in the sentiment and not giving biased category. The nature of the distribution is consistent with the methodology of the financial text analysis used by Onozo et al. [17] when continuous sentiment scoring is used as a more informative tool than categorical classifications.

5. Conclusion

The study has conducted a systematic study of the effectiveness of using news sentiment based on LLM and historical market data to predict stock movement by using a detailed multi-source fusion model. The results of the experiment, depicted in the visual model in the figure 2 titled Model Performance Comparison reveal quantifiable positive changes in predictive accuracy with sentence integration and this supports the main thesis based on References [1], [4] and [15]. The proposed XGBoost with Sentiment model had an accuracy of 47.72% that is an improvement of 2.77per cent over the average baseline performance of 46.43. Although this enhancement may seem minor, in the absolute sense, the implications it has are substantial, considering the limitations of the efficient market assumption and the need to overcome the barriers to achieving recurrent and steady outperforming the market standards in the past.

The analysis of feature importance shows that sentiment scores became one of the most important predictors in the proposed model, which confirms the theoretical premises made by Snasel et al. [1] on the importance of multi-source information fusion. This observation is consistent with the sentiment extraction improvements that have been reported by Onozo and others [17] and Chen and Kawashima [26], and it proves that the sentiment that has been extracted using LM gives relevant market information that cannot be adequately read using conventional technical indicators. The low correlation coefficients between the sentiment features and price-based metrics means that their relationship is orthogonal which is further support of the complementary information hypothesis of multi-mode approaches. The applied implementation in Google Colaboratory manages to overcome the issues of accessibility raised in intricate agent-based systems [8], [10] to offer computationally efficient framework preserving the advanced analytical attributes of transformer-based LLMs. This application shows that meaningful sentiment mixture is possible without having to consume a lot of computational resources or have special infrastructure and that advanced prediction techniques are affordable to larger research and practitioner communities.

A number of important lessons can be learned out of the overall assessment. To begin with, the statistical accuracy measure is similar in both the simulated trading situations, but the improved economic performance does indicate that the effect of sentiment integration is to achieve greater returns risk-adjusted by obtaining greater prediction confidence and stability. Secondly, the robustness of the model is demonstrated by the stable performance in various market conditions and compared to the single-source strategies especially in the sentiment-driven market regimes. Third, the effective submission to Indian equity data makes the sentiment-based prediction methods geographically valid in more than the traditionally analyzed markets in the West. Nonetheless, the study also finds out that there are still some challenges in prediction of financial markets. The fact that all models ganged together close to the random classifier performance in ROC analysis is indicative of the inherent challenges faced in winning in the market by using computational methods. This observation is consistent with the constraints as articulated by Elahi and Taghvaei [4] and highlights the fact that there should be restraint on expectation on the accurate prediction in highly efficient markets.

The research makes a number of important contributions to financial informatics. In terms of methodology, it also offers a reproducible system of implementing the LLM-based sentiment extraction and multi-source fusion. It empirically provides a complete validation of the feature of sentiment in prediction models. In real practice it illustrates the possibility of sophisticated methods of analysis in the available computational settings. In theory, it transforms the multi-source fusion ideas [1], [12] to modern text processing based on LLM models.

Some of the promising avenues should be examined on future research directions. To improve sentiment extraction accuracy, first, the domain-specific fine-tuning of LLMs with the help of financial corpora, as proposed by Ni et al. [11], might be useful. Second, the extension of the multi-source framework to cover other data forms like social media sentiment, earnings call transcript, and macroeconomic indicators may have complementary predictive information. Third, the model can be made more responsive to regime shifts by exploring the dynamic fusion techniques that adjust the weights of various information sources depending on market conditions and following the footsteps of Yue [7]. Lastly, an increase in the temporal resolution to intraday analysis could result in the identification of more short-term sentiment incorporation patterns in the price action.

To sum up, the study offers qualified yet valuable evidence that confirms the implementation of sentiment that is extracted by the LLM in stock movement prediction models. Though the absolute improvement in performance is limited by market efficiency, the improvements in feature significance, economic performance, and prediction stability are shown, which justifies the practical use of multi-source fusion strategies. The research forms a convenient basis to further studies as well as being part of the developing knowledge on information integration in financial markets. The results indicate that sentiment analysis is not sufficient to break the market efficiency; still, its strategy application to conventional indicators offers quantifiable improvements on prediction systems, especially in risk-managed practical use, where the long-term incremental improvements are meaningful.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

DATA AVAILABILITY STATEMENT

The authors confirm that the data supporting the findings of this study are available within the article [and/or] its supplementary materials. Code link: Google Colaboratory

AUTHOR CONTRIBUTIONS

• Anandkumar Pardeshi: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Investigation, Visualization, Writing – Original Draft Preparation.

• Dr. Sujata Deshmukh: Supervision, Validation, Resources, Writing – Review & Editing.

REFERENCES

[1] V. Snášel, J. D. Velásquez, M. Pant, D. Georgiou, and L. Kong, "A generalization of multi-source fusion-based framework to stock selection," Information Fusion, vol. 102, p. 102018, Feb. 2024, doi: 10.1016/j.inffus.2023.102018.

[2] N. Mishra, M. Anzar, S. Pandey, and S. Mishra, "A Review on Stock Market Trends and Stocks Price Prediction Using Sentiment Analysis and Market Data," in 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI), Greater Noida, India: IEEE, Apr. 2025, pp. 56–62. doi: 10.1109/ICCSAI64074.2025.11063888.

[3] G. Mu, N. Gao, Y. Wang, and L. Dai, "A Stock Price Prediction Model Based on Inves-tor Sentiment and Optimized Deep Learning," IEEE Access, vol. 11, pp. 51353–51367, 2023, doi: 10.1109/ACCESS.2023.3278790.

[4] A. Elahi and F. Taghvaei, "Combining Financial Data and News Articles for Stock Price Movement Prediction Using Large Language Models," in 2024 IEEE International Con-ference on Big Data (BigData), Washington, DC, USA: IEEE, Dec. 2024, pp. 4875–4883. doi: 10.1109/BigData62323.2024.10825449.

[5] M. Abdelsamie and H. Wang, "Comparative Analysis of LLM-based Market Prediction and Human Expertise with Sentiment Analysis and Machine Learning Integration," in 2024 7th International Conference on Data Science and Information Technology (DSIT), Nanjing, China: IEEE, Dec. 2024, pp. 1–6. doi: 10.1109/DSIT61374.2024.10881868.

[6] A. Pardeshi and S. Deshmukh, "Deep Learning in Stock Market Forecasting: Compara-tive Insights and Future Directions," in 2025 International Conference on Emerging Trends in Industry 4.0 Technologies

(ICETI4T), Navi Mumbai, India: IEEE, June 2025, pp. 1–6. doi: 10.1109/ICETI4T63625.2025.11132153.

[7] D. Yue, "Dynamic Weighted Multimodal Financial Forecasting Models: Fusion Strate-gies and Market Validation," in 2025 4th International Conference on Artificial Intelli-gence, Internet and Digital Economy (ICAID), Guangzhou, China: IEEE, Apr. 2025, pp. 60–66. doi: 10.1109/ICAID65275.2025.11034630.

[8] S. Chandra and G. Balakrishna, "Enhancing FinRL Trading Agents with Advance LLM-Processed Financial News: An Improved Approach Using DeepSeek-V3," in 2025 IEEE 11th International Conference on Intelligent Data and Security (IDS), New York City, NY, USA: IEEE, May 2025, pp. 52–54. doi: 10.1109/IDS66066.2025.00016.

[9] D.-Y. Zhou et al., "Federated Experiments: Generative Causal Inference Powered by LLM-Based Agents Simulation and RAG-Based Domain Docking," IEEE/CAA J. Au-tom. Sinica, vol. 12, no. 7, pp. 1301–1304, July 2025, doi: 10.1109/JAS.2024.124671.

[10] Y. Yu et al., "FinMem: A Performance-Enhanced LLM Trading Agent With Layered Memory and Character Design," IEEE Trans. Big Data, pp. 1–18, 2025, doi: 10.1109/TBDATA.2025.3593370.

[11] H. Ni et al., "Harnessing Earnings Reports for Stock Predictions: A QLoRA-Enhanced LLM Approach," in 2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS), Hangzhou, China: IEEE, Aug. 2024, pp. 909–915. doi: 10.1109/DOCS63458.2024.10704454.

[12] Z. Li, "Incorporating related stock and text for stock price movement prediction based on information fusion," in 2025 6th International Conference on Computing, Networks and Internet of Things (CNIOT), Shanghai, China: IEEE, May 2025, pp. 1–6. doi: 10.1109/CNIOT65435.2025.11070984.

[13] N. Madhuri, R. Tamilkodi, M. S. Teja, M. Arjun Kumar, N. R. Swaroop, and L. Jaya-prakash, "Investor Sentiment based Predictive Stock Market Forecasting using Deep Learning," in 2024 5th International Conference on Data Intelligence and Cognitive In-formatics (ICDICI), Tirunelveli, India: IEEE, Nov. 2024, pp. 1264–1269. doi: 10.1109/ICDICI62993.2024.10810985.

[14] X. Lu, J. Qiu, Y. Yang, C. Zhang, J. Lin, and S. An, "Large Language Model-Based Bidding Behavior Agent and Market Sentiment Agent-Assisted Electricity Price Predic-tion," IEEE Trans. on Enrgy Mrkts, Pol and Reg, vol. 3, no. 2, pp. 223–235, June 2025, doi: 10.1109/TEMPR.2024.3518624.

[15] C. Liu, A. Arulappan, R. Naha, A. Mahanti, J. Kamruzzaman, and I.-H. Ra, "Large Language Models and Sentiment Analysis in Financial Markets: A Review, Datasets, and Case Study," IEEE Access, vol. 12, pp. 134041–134061, 2024, doi: 10.1109/ACCESS.2024.3445413.

[16] Y. Abe, S. Matsuo, R. Kondo, and R. Hisano, "Leveraging Large Language Models for Institutional Portfolio Management: Persona-Based Ensembles," in 2024 IEEE Interna-tional Conference on Big Data (BigData), Washington, DC, USA: IEEE, Dec. 2024, pp. 4799–4808. doi: 10.1109/BigData62323.2024.10825362.

[17] L. Réka Ónozó, F. Viktor Arthur, and B. Gyires-Tóth, "Leveraging LLMs for Financial News Analysis and Macroeconomic Indicator Nowcasting," IEEE Access, vol. 12, pp. 160529–160547, 2024, doi: 10.1109/ACCESS.2024.3488363.

[18] L. Zhou, Y. Chen, J. Yu, S. Yongchareon, J. Ruan, and Y. Zhang, "LLM-Augmented Enhanced Graph Transformer for Stock Movement Prediction," in 2025 IEEE Region 10 Symposium (TENSYMP), Christchurch, New Zealand: IEEE, July 2025, pp. 1–8. doi: 10.1109/TENSYMP63728.2025.11144933.

[19] S. H. Saffarian and S. Haratizadeh, "LLM-Driven Feature Extraction for Stock Market Prediction: A Case Study of Tehran Stock Exchange," in 2024 15th International Con-ference on Information and Knowledge Technology (IKT), Isfahan, Iran, Islamic Re-public of: IEEE, Dec. 2024, pp. 59–65. doi: 10.1109/IKT65497.2024.10892779.

[20] Q.-Q. He, P. C.-I. Pang, and Y.-W. Si, "Multi-source Transfer Learning with Ensemble for Financial Time Series Forecasting," in 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Mel-bourne, Australia: IEEE, Dec. 2020, pp. 227–233. doi: 10.1109/WIIAT50758.2020.00034.

[21] Y. Hu, "Optimizing Economic Decision-Making Models through Multimodal Data Pro-cessing in Intelligent Systems," in 2025 4th International Conference on Artificial Intel-ligence, Internet and Digital Economy (ICAID), Guangzhou, China: IEEE, Apr. 2025, pp. 123–126. doi: 10.1109/ICAID65275.2025.11034649.

[22] B. N. Xia, A. Gupta, and M. J. Zaki, "Semantic Graph Learning for Trend Prediction from Long Financial Documents," in 2024 IEEE Symposium on Computational Intelli-gence for Financial

Engineering and Economics (CIFEr), Hoboken, NJ, USA: IEEE, Oct. 2024, pp. 1–8. doi: 10.1109/CIFEr62890.2024.10772910.

[23] L. Shang, “Sentiment-Driven Bitcoin Price Range Forecasting: Enhancing CART Deci-sion Trees With High-Dimensional Indicators and Twitter Dynamics,” IEEE Access, vol. 13, pp. 60508–60518, 2025, doi: 10.1109/ACCESS.2025.3557186.

[24] J. Wang and Z. Chen, “SPCM: A Machine Learning Approach for Sentiment-Based Stock Recommendation System,” IEEE Access, vol. 12, pp. 14116–14129, 2024, doi: 10.1109/ACCESS.2024.3357114.

[25] L. Cheng, J. Wen, and Y. Wang, “Stock Prediction Model Based on Multi-feature Graph Attention Network,” in 2025 8th International Conference on Artificial Intelli-gence and Big Data (ICAIBD), Chengdu, China: IEEE, May 2025, pp. 157–162. doi: 10.1109/ICAIBD64986.2025.11081993.

[26] Q. Chen and H. Kawashima, “Stock Price Prediction Using LLM-Based Sentiment Analysis,” in 2024 IEEE International Conference on Big Data (BigData), Washington, DC, USA: IEEE, Dec. 2024, pp. 4846–4853. doi: 10.1109/BigData62323.2024.10825946.