Enhancing Stock Market Trend Prediction Using Explainable Artificial Intelligence and Multi-source Data

John Ranjith[1], Kumar Chandar S[2,*]

[1]Research Scholar, School of Business and Management, CHRIST (Deemed to be University), Bengaluru, Karnataka, India

[2]Professor, School of Business and Management, CHRIST (Deemed to be University), Bengaluru, Karnataka, India

Emails: john.r@res.christuniversity.in; kumar.chandar@christuniversity.in

Abstract

Determining the trend of the stock market is a complex task influenced by numerous factors like fundamental variables, company performance, investor behavior, sentiments expressed in social media, etc. Although machine learning models support predicting stock market trends using historical or social media data, reliance on a single data source poses a serious challenge. This study introduces a novel Explainable artificial intelligence (XAI) to address a binary classification problem wherein the objective is to predict the trend of the stock market, utilizing an integration of multiple data sources. The dataset includes trading data, news and Twitter sentiment, and technical indicators. Sentiment analysis and the Natural Language Toolkit are utilized to extract the qualitative information from social media data. Technical indicators, or quantitative characteristics, are therefore generated from trade data. The technical indicators are fused with the stock sentiment features to predict the future stock market trend. Finally, a machine learning model is employed for upward or downward stock trend predictions. The proposed model in this study incorporates XAI to interpret the results. The presented model is evaluated using five bank stocks, and the results are promising, outperforming other models by reporting a mean accuracy of 90.14%. Additionally, the proposed model is explainable, exposing the rationale behind the classifier and furnishing a complete set of interpretations for the attained outcomes.

Keywords: Explainable artificial intelligence; Machine learning; Stock market prediction; Multi-source data; Sentiment analysis; Technical indicators

1. Introduction

The stock market plays a crucial role in the economy by promoting financial growth. A stock market is a platform for economic transactions for the selling and buying of shares. These shares represent ownership stakes in businesses, either through equity or market ownership. They can range from public market shares to privately traded shares of companies [1]. Funds from small investors are transferred to larger trading institutions, including banks and businesses, by means of stock market transactions. However, given the dynamic and unpredictable nature of the stock market, investment in it is seen as very risky.

Through dividends provided by the company's shareholder incentive scheme, investors may purchase additional stocks and make substantial profits. They can also engage in stock trading with other market participants through electronic platforms and stock brokerages. Stock traders aim to purchase stocks expected to rise in value while selling those expected to fall [2]. Consequently, stock traders must be able to predict stock movements before deciding to buy or sell. The more precise their predictions regarding stock movement, the greater their potential profit [3],[4]. Thus, creating an algorithm capable of accurately predicting stock market behavior becomes crucial to assist traders in maximizing profits. However, predicting stock market behavior poses challenges due to multiple factors.

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

Numerous approaches have been designed for predicting stock trends. These approaches can be categorized into four types: the first approach is fundamental analysis, which relies on public financial data [5],[6], the second approach involves technical analysis which uses past data for making recommendations [7],[8], the third approach incorporates Machine learning (ML) techniques to process the enormous volume of data obtained from different sources [9], and the fourth type is Sentiment analysis (SA) [36], which predicts trend by evaluating published news and blogs [10],[11]. However, as compared to employing several sources of data, using one source of data may not provide the best results. The fusion of diverse data is relatively more recent compared to single data, and research indicates that they could have a more significant impact on the decision to sell or buy stocks [12]. Hence, this study provides a strategy for predicting stock movement utilizing multi-source data fusion such as financial data, news and Twitter sentiment, and technical indicators.

While ML models offer superior predictive capabilities, they suffer from complexity and a lack of interpretability [13]. However, the need for comprehension and explanation is essential for internal applications within fields like investment, banking, and customer service. Explainable artificial intelligence (XAI) [14],[15] emerges as a solution to address this issue. Although standards for XAI are still in the process of development, ongoing trends indicate that XAI will become an important requirement for future economic Artificial intelligence (AI) uses. Based on the information from the Gartner website, it is projected that by 2050, 30% of major businesses and government contracts for procuring AI uses will necessitate AI systems that are both explainable and ethically sound [16]. This study proposes an XAI framework using multi-source data for stock trend prediction. The major innovations and scientific contributions of this study are as follows:

(i) An innovative architecture XAI is built for predicting stock market trends by using multi-source data fusion techniques.
(ii) A comprehensive dataset incorporating historical stock data, news sentiment, Twitter sentiment, and technical indicators is constructed to improve the efficiency and robustness of stock price trend predictions.
(iii) The effectiveness and reliability of the proposed framework have been validated through rigorous testing on five bank sectors, showcasing its potential for accurate stock trend prediction.
(iv) A detailed examination of the experimental outcomes attained from applying the presented framework to the selected bank sectors, including performance metrics and comparative evaluations.
(v) The impact and significance of integrating multi-source data fusion with XAI are investigated to increase the accuracy and interpretability of stock market trend predictions.
(vi) The proposed model aimed to exploit the valuable insights collected from multi-source data to improve the accuracy of stock price predictions, thereby bridging the gap between multi-source data and financial markets.
(vii) Innovative integration of multi-source data fusion with XAI is done specifically to predict stock market trends, making a notable advancement in the fields of financial prediction and algorithmic testing.

The remaining sections of the paper are structured as follows. An overview of the existing literature on stock market prediction is provided in Section 2. Section 3 describes the methodology proposed. The results of the experiments are elaborated in Section 4. The study concludes with concluding remarks in Section 5.

2. Related Works

ML models are well known for their ability to predict stock market movements. This section reviews previous efforts to predict stock market movements using various strategies. Models that use technical analysis principles for stock trend prediction treat it as a classification problem, wherein they use previous data to understand market tendencies. In the past, prior to the widespread availability of financial news and social media data, and techniques for text data scoring, researchers adopted different ML techniques on past data to make predictions. Individuals may express their experiences and opinions on a range of subjects through social media platforms. Financial news, social media activities, and political news directly impact the stock market. With all this data accessible online, integrating it with past data and using ML techniques has substantially improved prediction accuracy and the trend of prediction [12],[13].

Lee [1] used a subset of the NASDAQ Index from the Taiwan Economic Journal Database (TEJD) from 2008 and a hybrid feature selection approach in combination with a Support vector machine (SVM) for predicting movements in the stock market. The feature selection involved a hybrid technique, utilizing F-score and sequential search forward. The primary weakness in the model is that it did not compare the SVM's performance to that of other methods, instead comparing it to the Back propagation neural network (BPNN).

Kara et al. [2] implemented SVM and Artificial neural network (ANN) to accomplish stock trend prediction tasks based on historical data. Historical data spanning from 1997 to 2007 was gathered from the Istanbul Stock Exchange (ISE) for analysis. However, the limitation of this study was that the analysis was restricted to ISE

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

stocks. This drawback raises concern about the model's generalizability to other stock exchanges, necessitating further validation. Ni et al. [3] employed SVM to anticipate the trend of stock prices. The performance of the SVM was analyzed using data consisting of 19 technical indicators from the Shanghai Stock Exchange Composite Index (SSECI). Fractal feature selection was utilized to fine-tune the model. The model showed promising outcomes. However, the model was tested with only technical indicators and overlooked macro factors. Lei [4] adopted a Wavelet neural network (WNN) to predict stock price trends. A rough set was used to select attributes and configure the WNN structure. Stock market trends were predicted based on historical data.

Thakur & Kumar [5] devised a hybrid model to forecast stock trends using Random forest (RF) in combination with multi-class classifiers. Stocks from three prominent markets such as NASDAQ, S&P 500, NIFTY Bank, and NIFTY 50, were used for analysis. The hybrid framework fused the power RF with SVM to generate signals indicating Buy or Hold or Sell signals. However, the model relied solely on historical data, omitting other factors such as technical indicators and text data. Qiu & Song [6] designed an optimized model to predict the movement of the Japanese Stock Market (JSE) by fusing ANN with a Genetic algorithm (GA). The integration of GA was aimed at optimizing the parameters of the ANN to enhance its predictive capabilities. A dataset was used, covering the period from 2007 to 2013. However, the proposed model reported low prediction accuracy.

Long et al. [8] devised a model to predict stock market movements using a multifilter neural network. The network was designed by combining convolutional and recurrent neurons. In this model, four factors including open, high, low, and close prices were utilized as input for the network to make predictions. The model was tested on stock data from the Chinese stock market index. Results proved the model's potential to capture stock market trends. This model requires a larger volume of training data to generate accurate predictions. Khedr et al. [10] predicted stock market trends using ML techniques based on Twitter data. Twitter data was collected, preprocessed, and then computed sentiment scores to measure the overall sentiment about various stocks. RF was adopted to predict stock trends. Nonetheless, the RF performance suffered due to its dependence on Twitter data. Combining Twitter sentiment score with quantitative data might enhance prediction accuracy.

Mehta et al. [11] presented a comparative study by exploring the strength of several classifiers such as SVM, Linear Regression (LR), and Long Short-Term Memory (LSTM) in anticipating stock market trends. Historical data was combined with news sentiment scores to increase the prediction rate. Experimental findings showed that the combined data gave better results. Vargas et al. [12] used a deep learning algorithm to design a stock prediction model. Primarily, two types of data such as historical data and news data were gathered. Secondly, these data were preprocessed and then technical indicators and sentiment scores were computed. Finally, a Convolutional neural network (CNN) performed the prediction. Results revealed that combining quantitative data with qualitative data can improve the prediction result. However, the model necessitates additional training data to get accurate predictions.

Kabbani & Usta [27] introduced a system for forecasting future stock trends employing ML classifiers. In this system, technical indicators were derived from historical data, and sentiment scores were computed from news data. The indicators were merged with sentiment scores to form a unified feature matrix. Three ML classifiers including LR, RF, and gradient boosting machines were adopted to forecast future stock movement. Among these models, RF emerged as the top performer by reporting an accuracy of 63.58%. This approach proved that qualitative data can be integrated with quantitative data indicators to increase prediction rates. Xiao & Ihnaini [28] predicted stock market trends using ML and a fusion of Twitter data, news data, and historical prices. These datasets were gathered and processed to extract features. Diverse ML classifiers, including SVM, RF, LR, and Naïve Bayes (NB) were employed to predict future stock trends. NB gave better outcomes than other classifiers. However, this model showed low accuracy.

Li et al. [29] implemented Graph neural networks (GNNs) to anticipate stock market movements. The proposed approach used various types of data including social media data, historical price, and technical indicators to anticipate stock price trends. The authors attempted to enhance prediction accuracy with the help of GNNs. Experimental outcomes demonstrated excellent performance. However, the training process demands more time and computational resources due to the complex nature of GNNS and the large volume of data involved. Theissler et al. [30] provided a comprehensive review and research directions on the application of XAI for time series prediction. The authors explored previous methods, classified different methods, and identified key research directions. Additionally, they discussed diverse approaches used in XAI for time series prediction.

Maqbool et al. [31] introduced a stock prediction model using ML. The authors focused on fusing sentiment scores computed from news data with Multilayer perceptron (MLP). Through empirical evaluation, they demonstrated the power of their approach in the prediction of stock prices. Although sentiment analysis can provide valuable insights, it may not capture all relevant factors influencing stock market trends.

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

Koukaras et al. [32] presented a study on stock market prediction utilizing microblogging sentiment analysis combined with ML techniques. Their research explored the effectiveness of incorporating sentiment analysis of Twitter data, as a predictor for stock market movements. However, the main drawback of this method could be the inherent noise and volatility of social media sentiment data. Twitter data is susceptible to manipulation, which may introduce inaccuracies in sentiment analysis and consequently affect the prediction accuracy. Chen et al. [33] devised a model fusion approach for predicting stock market trends using multi-source data. The model incorporated weighted unstructured and structured data from multiple sources to predict future stock market movements. Various ML algorithms such as RF, LSTM, and MLP were used to predict future movements. However, this approach does not address potential challenges related to data quality or the dynamic nature of financial markets, which could impact the model's performance.

Li et al. [34] examined theoretical models employed at various levels of data fusion-data level, feature level, and decision level fusion in the context of stock price prediction. By reviewing the evolution of stock price prediction from a data fusion perception, their research highlights the widespread and effective utilization of data fusion methods in this domain. In addition to this, their investigation provided future scope for research in this field. Ma et al. [35] created a multisource aggregated classification model to predict stock market behavior. In this approach, historical data, market-driven news sentiments of target stocks along with the news sentiments of related stocks were merged. To enhance the depiction of market sentiments from news data, authors pre-trained a feature generator to fit news to actual trends of the stock market. This method incorporated a Graph convolutional network (GCN) to capture the impact of news from related companies on the target stock. By using these features, this approach provided better results. However, this approach may involve high computational complexity associated when dealing with larger datasets.

3. Proposed Stock Market Movement Prediction System

To predict stock movement, the primary objective of this study is to examine the viability of employing XAI. Figure 1 shows the general articulation of the presented framework. The presented XAI comprises the following phases: data preprocessing, feature extraction, prediction, and explanation.

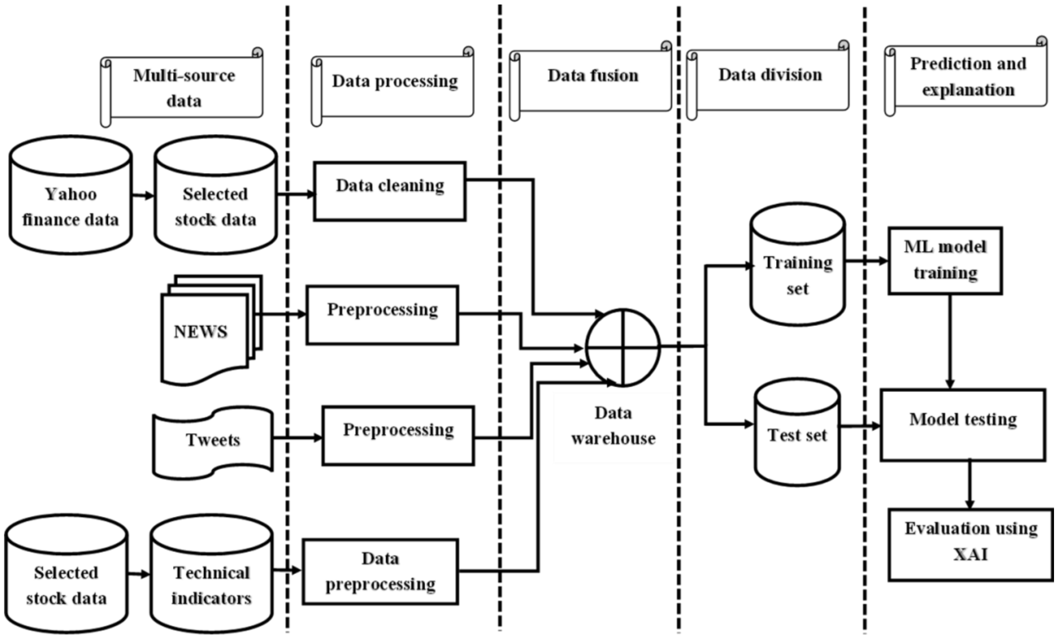


Figure 1. The general architecture of the proposed stock movement prediction system

3.1. Statement of the problem

The input for the proposed system comprises five banking stocks, including Axis Bank Limited, HDFC Bank Limited, ICICI Bank Limited, IndusInd Bank Limited, and Kotak Mahindra Bank Limited. These stocks include trading data within the Indian stock market, a compilation of news articles from credible press sources, and data from Twitter. The significant objective of this study is to predict the stock market trend on a specified day for a specific bank stock.

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

3.2. Data preprocessing

Social media platforms serve as the primary networking source for public companies, governments, and businesses to share their thoughts and information. The social media data is chosen due to its conciseness. After data gathering, a series of preprocessing steps are done, as displayed in Figure 2 to refine social media data.

Standard Natural Language Processing (NLP) pipeline is used to preprocess social media data. Tokenization, stop word elimination, stemming, and Lemmatization is carried out using the Natural Language Toolkit (NLTK) [17],[18] library in Python 3. Tokenization is the first step, dividing the text into individual tokens, and facilitating easier analysis by isolating distinct words from the collected data. Stop words, HTML tags (<, >), punctuation marks (commas, semicolons, spaces), and URLs are removed due to their lack of valuable sentiment-carrying information. Words are reduced to their basic forms by stemming and lemmatization. Lemmatization condenses words to their basic form based on their parts of speech, while stemming eliminates suffixes like "ed" or "ing" to simplify words to their base form.

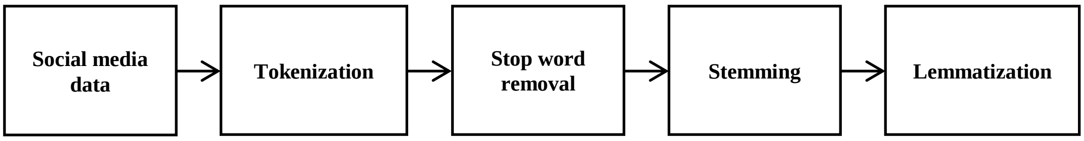


Figure 2. Social media data preprocessing

3.3. Computation of Technical indicators

Simple moving average (SMA) and exponential moving average (EMA) are the two technical indicators taken into consideration in this analysis. These technical indicators are computed as follows,

$$\mathrm{SMA} = \frac{\mathrm{C_1+C_2+C_3 \ldots . C_n}}{\mathrm{n}} \tag{1}$$

$$\mathrm{EMA} = c_t * \frac{2}{1+N} + PREVIOUS\ EMA * (1 - \frac{2}{1+N}) \tag{2}$$

Using Equations (1) and (2), SMA10, SMA20, EMA10, and EMA20 are computed. Following the computation of technical indicators, the preprocessed data is combined to create a unified feature matrix.

3.4. Prediction

The feature vectors obtained are used as input for the predictor. Wang et al. [19] showed that ML models achieved promising results to other methods in time series classification. This study served as inspiration for the construction of SVM [20],[37], which is used to do analysis and anticipate stock trends. SVM is built upon the concept of defining hyperplanes that separate data points within high-dimension feature spaces. The linear SVM model concentrates on optimizing the separation margin among positive and negative hyperplanes. The classification procedure can be expressed as,

$$y_i = \begin{cases} +1 & if\ b + \propto^T x \geq +1 \\ -1 & if\ b + \propto^T x \geq -1 \end{cases} \tag{3}$$

The characteristics are transformed into a higher-dimensional space using a kernel function for non-linear situations. In this investigation, the Radial basis function (RBF) is used as kernel, which is defined in Equation (4),

$$Kernel(x, y) = exp(-\gamma \|x - y\|^2) \tag{4}$$

3.5. Explainable Artificial Intelligence (XAI)

There has been a growing interest in applying artificial intelligence in executing financial tasks like algorithmic trading, and so on. In Particular, machine learning models have captured the attention of researchers aiming to build a smarter model for prediction and classification than conventional models. Artificial intelligence-based models have shown superior results due to their ability to handle unbalanced and complex data. This advantage makes them more accurate in analyzing financial data. Additionally, a new concept in artificial intelligence enables machine learning models to interpret their results of complex data. This enables humans to not only find solutions

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

to complex data but also understand and interpret the reason behind the obtained solutions. This new concept is classically referred to as Explainable Artificial Intelligence (XAI). As shown in Figure 3, the explainer presents the model's interpretation to the user so they may make an informed choice after the model's training and prediction-making. The functioning of XAI involves the model generating a prediction, together with the data, undergoing analysis using the XAI tool. This tool describes the data by considering biases, aiming to improve the decision-making processes. After studying various XAI [21],[38] tools like what-if, LIME [22], and Shapely [23]-[26], this investigation makes use of LIME. The predictive model's behavior and the attributes it chooses to predict are understood using LIME. Lime is local which means that it explains every single observation that the data have. It is mode-agnostic which means that any black-box mode can be developed in the future. The algorithmic steps of the proposed method are given in Table 1.

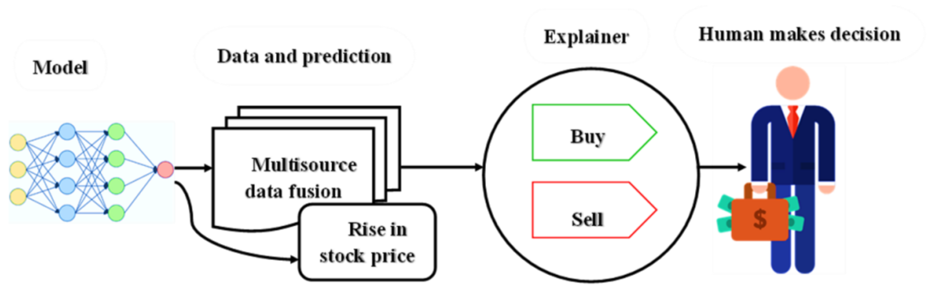


Figure 3. Functioning of XAI

Table 1: Algorithmic processes of the proposed method

Input: historical data, h, news sentiment, NS, Twitter sentiment, TS, technical indicators, TI
Output: Stock price trend
Step 1: Read historical data, news data, and Twitter data from the database
Step 2: Assign labels for stock prices
Step 3: Preprocess news and Twitter data and then compute NS and TS
Step 4: Compute TI from H using Equation (1) and Equation (2)
Step 5: Create a multisource data comprising historical data, news sentiment, Twitter sentiment, and technical indicators: MD = {h, NS, TS, TI}
Step 6: Split the data into training, validation, and testing data
Step 7: Train the SVM using training data and optimize its parameters
Step 8: Test the model using unseen data
Step 9: Compare the model's output with the actual labels
Step 10: Interpret the model via Local Interpretable Model-agnostic Explanations (LIME)

4. Numerical Results and Discussion

This section assesses the effectiveness of the developed XAI in predicting stock market trends utilizing data from multiple sources. The presented system's design is implemented using MATLAB 2022 on an 11th gen Intel® Core ™ i5 processor @2.70GHz with 16GB of RAM. Python is employed to preprocess social data.

4.1. Model performance evaluation

Four metrics described below are used to assess the model's effectiveness:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

$$F1 - score = 2 * \frac{(Precision*Recall)}{(Precision+Recall)} \quad (8)$$

where the variables true positive, true negative, false positive, and false negative are, respectively, TP, TN, FP, and FN.

4.2. Dataset description

The primary data source used in this investigation comprises historical data collected from five banks listed in the Indian stock market. For each bank, data is gathered at a daily frequency, including information like the opening price, lowest price, highest price, closing price, and volume. The second source data includes the sentiments expressed in news articles. The third data source pertains to sentiments conveyed on the Twitter platform, while technical indicators constitute the fourth. Table 2 offers a summarized overview of the data used in this study.

Table 2: Details of data

Number of stocks	Five
Sector	Banking
Quantitative data	Trading data and technical indicators
Qualitative data	News and Twitter sentiment score
Period	01/01/2018-29/09/2023
Total number of samples	1422
Training period	01/01/2018-11/01/2022
Validation period	12/01/2022-05/08/2022
Test period	08/08/2022-29/09/2023
Classes	2 (rise/fall)

4.3. Prediction results and discussions

The aim of this experimental assessment has three primary objectives:

(i) To demonstrate that the fusion of multiple data sources and the predictive method introduced in this study effectively captures the relationship between the fusion of multi-source data and the trends in stock market prices.

(ii) To illustrate that the SVM classifier can generate interpretable explanations while maintaining performance metrics analogous to other state-of-the-art ML classifiers.

(iii) Evaluate the effectiveness of the proposed model with earlier approaches.

To determine the efficiency of the presented model in the context of stock market data analysis, several experiments are conducted in this study. To facilitate the evaluation, the entire data is partitioned into three subsets: 10% percent of the data is used for validation, 70% percent is used for training machine learning classifiers, and the remaining portion is used to assess how well the learned classifiers predict the future. Furthermore, the threshold that determines the class is set empirically to a 2% daily variation, representing a significant change in stock trends. Table 3 presents the hyperparameter configurations and model evaluations concerning performance metrics.

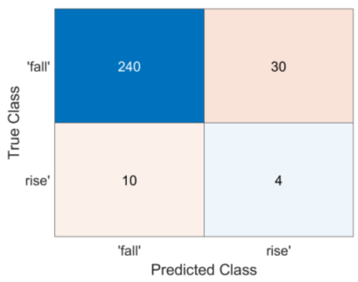
It is noted that accuracy alone may not provide a complete and accurate assessment of the proposed model's performance, particularly when dealing with imbalanced datasets. Accuracy fails to account for the relative importance of FP over FN. Hence, recall and precision are emphasized as more reliable measures for evaluating model performance, especially in scenarios of imbalanced datasets. Furthermore, this study incorporates the F1-score, a metric that signifies the harmonic mean of recall and precision, to provide a comprehensive assessment. Table 3 presents the prediction outcomes of the proposed system for varying data types, specifically quantitative data, qualitative data, and multi-source data. The key findings derived from the outcomes can be summarized as follows:

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

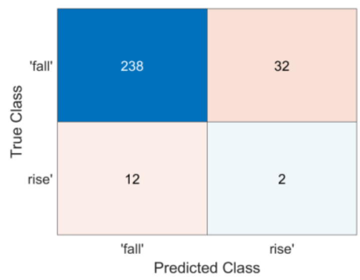
(i) Classification performance: A higher accuracy means the model can accurately predict a significant amount of data of both rising and falling stock market movements within the stock market among all the predictions made. The proposed model demonstrated superior classification abilities for multi-source data either using quantitative or qualitative data alone, achieving the highest accuracy of 91.20% among the different input feature types, indicating its overall effectiveness in making accurate predictions.

Table 3: Prediction outcomes of the proposed system for varying input features

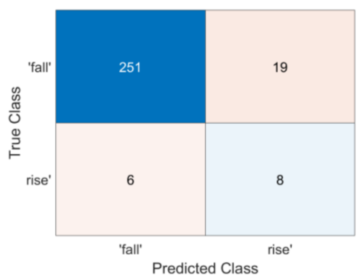
Data	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Quantitative data	85.92	88.89	96.00	92.31
Qualitative data	84.51	88.15	95.20	91.54
Proposed	91.20	92.96	97.67	95.26



(a)



(b)



(c)

Figure 4. Confusion matrices: (a) Quantitative data, (b) Qualitative data, and (c) Multi-source data

(ii) Precision and Recall importance: Precision and recall are essential in the context of imbalanced datasets, which is often the case when dealing with stock market data. Higher precision and recall measure the accuracy of positive predictions and the model's capability to find upward trends, respectively. The proposed model exhibited a high prediction of 92.96% for multisource input data, indicating a high percentage of correctly predicted positive cases out of all cases predicted as positive, along with a recall of 97.67%, signifying its remarkable accuracy in predicting positive instances.

(iii) F1-score and robustness: The F1-score is supposed to offer a fair assessment of precision and recall because the class distribution of rise and fall cases in stock data is unequal. The F1-score of 95.26% indicated a balanced performance between precision and recall, suggesting fewer FPs and FNs in the predictions accompanied by the proposed system.

The proposed model showed superior performance across all metrics compared to using either quantitative or qualitative data alone. It achieved higher precision, accuracy, recall, and F1-score, demonstrating its efficiency in making accurate predictions. Based on the analysis and outcomes presented in Table 3, it is proved that the developed XAI model has the potential to predict future trends in the stock market. Further to this, the confusion matrices of the XAI model for different input features are illustrated in Figure 4. These matrices highlight areas of strength and areas that require improvement in accurately predicting the classes, providing valuable guidance for refining the model's predictive capabilities.

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

4.4. Model Explainability

This section presents a complete qualitative assessment of the explainability of the presented stock price trend prediction model. The presented model utilizes the XAI technique named LIME to explain the interpretations of input features. This aids in understanding the impact of each attribute within the aggregated data on the prediction results of stock price movements. LIME is a powerful tool for decoding the outcome of ML algorithms. It provides valuable insights into their decision-making process. By computing LIME values, a clear understanding of the predictive model's internal operations is gained. Figure 5 depicts that five out of the six significant features are consistent for XAI. These features include Twitter emotions (x7), SMA (x4), highest price (x3), EMA (x10), lowest price (x2), and EMA (x10), thus reinforcing the reliability and robustness of the identified predictors within the proposed model. This finding highlights the consistency and stability of the identified predictors in the proposed model, further enhancing confidence in the model's predictive capabilities.

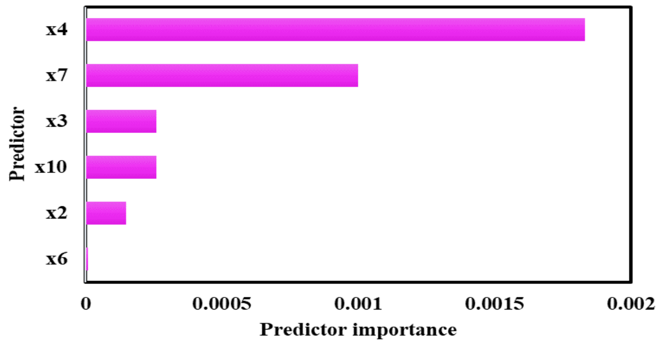


Figure 5. Significant features of XAI

4.5. Comparison with existing methods

The presented model is assessed with the other traditional models to prove its superiority in stock price trend prediction. These models are RF [27], MLP [31], LSTM [33], and GCN [35]. A comparative analysis is given in Table 4. The proposed model attained a maximum accuracy of 91.20%, indicating its superior ability to correctly predict stock market trends compared to previous approaches. Previous methods, RF, MLP, LSTM, and GCN have accuracies ranging from 79.93% to 83.80%, demonstrating lower predictive accuracy compared to the proposed model. The precision value for the proposed model is 92.96% and previous approaches range from 83.90% to 87.78%. The developed framework attained higher precision compared to previous methods, indicating its ability to correctly predict positive cases out of all cases predicted as positive. It also attained the highest recall of 97.67%, proving its capability to capture a large proportion of actual positive cases. Previous methods have recall values ranging from 94.56% to 95.04%. This range shows lower performance in capturing positive cases compared to the proposed model. The developed model attained an F1-score of 95.26% which indicates a balanced performance between precision and recall. Existing approaches have F1-score ranging from 88.80% to 91.15%. A lower F1-score signifies a comparatively lower balance between precision and recall compared to the proposed model. It is observed from the comparative analysis that the developed model exhibited superior performance to the existing approaches. A higher value of accuracy, precision, recall, and F1-score confirms the efficacy of the developed model in anticipating stock price movements with superior reliability and robustness. The higher F1-score of the proposed model further highlights its balanced performance in handling imbalanced datasets and making valuable predictions.

Table 4: Performance comparison between the proposed and existing approaches in stock trend prediction

Researchers	Predictor	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Kabbani & Usta [27]	RF	83.80	87.78	94.80	91.15
Maqbool et al. [31]	MLP	81.69	85.19	95.04	89.84
Chen et al. [33]	LSTM	80.99	84.44	95.00	89.41
Ma et al. [35]	GCN	79.93	83.70	94.56	88.80
Proposed model	XAI	91.20	92.96	97.67	95.26

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

4.6. Discussions on the Results

The application of XAI in anticipating stock market trends has been relatively underexplored in the literature, as most of the past models have been designed to address the challenges in forecasting stock market movements. This study sought to examine the impact of utilizing multiple sources of data in capturing stock market trends to fill the gap in the current research background. To address this issue, the study considered different types of data, including trading data, Twitter sentiments, news sentiments, and technical indicators. These data were gathered and then preprocessed to make them fit for further analysis. Subsequently, the SVM was designed and trained to predict future trends in the stock market. The findings confirmed that the combination of multi-source data significantly enhances the accuracy of predicting future stock market trends. The presented model provided excellent performance compared with other models. Particularly significant was the achievement of a testing accuracy of 90.14%, precision of 91.19%, recall of 97.94%, and F1-score of 94.44% in classifying stock market trends using multisource data fusion.

Figure 5 presents the interpretations of the effect of the features in stock trend prediction about the subject of interpreting the feature's influence in classification results. LIME was used to identify the most influential feature in stock market movement prediction. The model showed that the SMA emerged as the most important feature for detecting fluctuations in the stock market. This emphasizes the significance of certain features in identifying changes in the stock market.

This study showed the effectiveness of the proposed XAI model in anticipating stock market movements to provide valuable insights for investors and financial analysis. The higher values of precision, recall, F1-score, and accuracy highlight the robustness of the proposed model, particularly in imbalanced datasets. These results highlight the practical applicability of the proposed XAI model as a valuable tool in stock market analysis. In addition to this, this offers interpretable predictions and actionable insights for stakeholders in financial markets. In the future, continued exploration and refinement of XAI techniques hold promise for further increasing the effectiveness, interpretability, and applicability of stock price trend prediction models.

5. Conclusion and Future Scope

This study introduced an XAI model for predicting future trends and stock market movements based on multiple data fusions derived from various sources. Four different types of data such as trading data, news, and Twitter sentiments, and technical indicators are used to form a feature-rich matrix. Finally, an SVM classifier was employed to predict future stock market trends. Following the classification, LIME was used to analyze the influences of each feature on the predicted outcomes. The findings demonstrated that in comparison to ANN, DT, and RF, the presented XAI obtained greater accuracy, precision, recall, and F1-score, indicating its resilience and ability in stock market prediction. However, the proposed model has small limitations related to the sample size used for the analysis. In future research, there could be an emphasis on incorporating a wider range of technical indicators into the predictive model. Future research will concentrate on using Generative adversarial networks for processing textural information related to stock market sentiment and news. The effectiveness of the metaheuristic algorithms in combination with multi-source data will be investigated. Deep neural network models will be explored to improve prediction accuracy. Furthermore, other factors such as macroeconomic variables, and consumer behavior data will be considered to gain deeper insights into market dynamics and improve prediction performance.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Lee, M. C., Using support vector machine with a hybrid feature selection method to the stock trend prediction. Expert Systems with Applications, 36(8), 10896-10904, 2009.

[2] Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K., Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications, 38(5), 5311-5319, 2011.

[3] Ni, L. P., Ni, Z. W., & Gao, Y. Z., Stock trend prediction based on fractal feature selection and support vector machine. Expert Systems with Applications, 38(5), 5569-5576, 2011.

[4] Lei, L., Wavelet Neural Network Prediction Method of Stock Price Trend Based on Rough Set Attribute Reduction. Applied Soft Computing, 62, 923-932, 2018.

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

[5] Thakur, M., & Kumar, D., A hybrid financial trading support system using multi-category classifiers and random forest. Applied Soft Computing, 67, 337-349, 2018.

[6] Qiu, M., & Song, Y. (Eds.). (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. Public Library of Science (PLoS).

[7] Chen, L., Qiao, Z., Wang, M., Wang, C., Du, R., & Stanley, H. E., Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market? IEEE Access, 6, 48625-48633, 2018.

[8] Long, W., Lu, Z., & Cui, L., Deep learning-based feature engineering for stock price movement prediction. Knowledge-Based Systems, 164, 163-173, 2019.

[9] Billah, M., Waheed, S., & Hanifa, A., Stock market prediction using an improved training algorithm of neural network. International Conference on Electrical, Computer & Telecommunication Engineering, 2016.

[10] Khedr, A. E., Salama, S.E., & Yaseen, N., Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis. International Journal of Intelligent Systems and Applications, 9(7), 22-30, 2017.

[11] Mehta, P., Pandya, S., & Kotecha, K., Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. PeerJ Computer Science, 7, e476, 2021.

[12] Vargas, M. R., Dos Anjos, C. E. M., Bichara, G. L. G., & Evsukoff, A. G., Deep Leaming for Stock Market Prediction Using Technical Indicators and Financial News Articles. International Joint Conference on Neural Networks, 2018.

[13] Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S., Stock market prediction using machine learning classifiers and social media, news. Journal of Ambient Intelligence and Humanized Computing, 13(7), 3433-3456, 2020.

[14] Adadi, A., & Berrada, M., Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138-52160.

[15] Carta, S. M., Consoli, S., Piras, L., Podda, A. S., & Recupero, D. R., Explainable Machine Learning Exploiting News and Domain-Specific Lexicon for Stock Market Forecasting. IEEE Access, 9, 30193-30205, 2021.

[16] Kasey, P. Gartner, Top 10 strategic technology trends for 2020, available at http:// www.Gartner.Com/smarterwithgarther/gartner_top_10_strategic_technology_trends_for_2020.

[17] Team NLTK, NLTK Is a Suite of Libraries and Programs for Symbolic and Statistical Natural Language Processing for English Written in the Python Programming Language, 2001.

[18] Loper, E., & Bird, S., NLTK: The Natural Language Toolkit (Version 1). arXiv, 2001.

[19] Wang, Z., Yan, W., & Oates, T., Time series classification from scratch with deep neural networks: A strong baseline. International Joint Conference on Neural Networks, 2017.

[20] Tran, K. L., Le, H. A., Nguyen, T. H., & Nguyen, D. T., Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam. Data, 7(11), 160, 2022.

[21] Khaleghi B., The how of explainable AI: post-modelling explainability. Towards Media Science, 2019.

[22] Wexler, J., Pushkarna, M., Bolukbasi, T., Wattenberg, M., Viegas, F., & Wilson, J., The What-If Tool: Interactive Probing of Machine Learning Models. IEEE Transactions on Visualization and Computer Graphics, 2019.

[23] Messalas, A., Kanellopoulos, Y., & Makris, C., Model-Agnostic Interpretability with Shapley Values. International Conference on Information, Intelligence, Systems and Applications, 2019.

[24] Breiman, L. (Eds.). (2001). Machine Learning. Springer Science and Business Media LLC.

[25] Friedman, J. H. (Eds.). (2002). Stochastic gradient boosting. In Computational Statistics & Data Analysis. Elsevier BV.

[26] Ruck, D. W., Rogers, S. K., Kabrisky, M., Oxley, M. E., & Suter, B. W., The multilayer perceptron as an approximation to a Bayes optimal discriminant function. IEEE Transactions on Neural Networks, 1(4), 296-298, 1990.

[27] Kabbani, T., & Usta, F. E., Predicting The Stock Trend Using News Sentiment Analysis and Technical Indicators in Spark (Version 1). arXiv, 2022.

[28] Xiao, Q., & Ihnaini, B., Stock trend prediction using sentiment analysis. PeerJ Computer Science, 9, e1293, 2023.

[29] Li, X., Wang, J., Tan, J., Ji, S., & Jia, H., A graph neural network-based stock forecasting method utilizing multi-source heterogeneous data fusion. Multimedia Tools and Applications, 81(30), 43753-43775, 2022.

[30] Theissler, A., Spinnato, F., Schlegel, U., & Guidotti, R., Explainable AI for Time Series Classification: A Review, Taxonomy and Research Directions. IEEE Access, 10, 100700-100724, 2022.

[31] Maqbool, J., Aggarwal, P., Kaur, R., Mittal, A., & Ganaie, I. A., Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach. Procedia Computer Science, 218, 1067-1078, 2023.

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024

[32] Koukaras, P., Nousi, C., & Tjortjis, C., Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning. Telecom, 3(2), 358-378, 2022.

[33] Chen, X., Hirota, K., Dai, Y., & Jia, Z., A model fusion method based on multi-source heterogeneous data for stock trading signal prediction. Soft Computing, 27(10), 6587-6611, 2022.

[34] Li, A., Wei, Q., Shi, Y., & Liu, Z., Research on stock price prediction from a data fusion perspective. Data Science in Finance and Economics, 3(3), 230-250, 2023.

[35] Ma, Y., Mao, R., Lin, Q., Wu, P., & Cambria, E., Multi-source aggregated classification for stock price movement prediction. Information Fusion, 91, 515-528, 2023.

[36] Mohamed, L., Soliman, G., & Nasser, A., A Survey on Sentiment Analysis Algorithms and Techniques for Arabic Textual Data. Journal of Fusion: Practice and Applications, 2(2), 74-87, 2020.

[37] Shaymaa, A. A., & Amer, R., Detection and Classification of Alcoholics Using Electroencephalogram Signal and Support Vector Machine. Journal of Fusion: Practice and Applications, 2(1), 14-21, 2020.

[38] Abdelmonem, A. N., Interpretable Machine Learning Fusion and Data Analytics Models for Anomaly Detection. Journal of Fusion: Practice and Applications, 3(1), 54-69, 2021.

DOI: https://doi.org/10.54216/FPA.160211
Received: December 27, 2023 Revised: March 03, 2024 Accepted: June 09, 2024