Journal of Computer Science

Original Research Paper

# Stock Closing Price Forecasting Using LSTM, Sentiment Analysis, Kalman Filter

[1]Assel Abdildayeva, [2]Galymzhan Nurtugan, [3]Guldana Taganova and [2]Ardak Akhmetova

[1]*Artificial Intelligence and Big Data, Al-Farabi Kazakh National University, Kazakhstan*
[2]*Information Technology, Al-Farabi Kazakh National University, Kazakhstan*
[3]*Information Technology and Engineering, Astana International University, Kazakhstan*

*Article history*
Received: 07-02-2024
Revised: 15-02-2024
Accepted: 09-03-2024

Corresponding Author:
Assel Abdildayeva
Artificial Intelligence and Big Data, Al-Farabi Kazakh National University, Kazakhstan
Email: abass_81@mail.ru

**Abstract:** The article explores the problem of predicting the closing prices of stocks using an innovative approach that combines methods of machine learning, sentiment analysis and noise filtering. The focus is on the development and testing of a comprehensive model that integrates recurrent neural networks (LSTM) for time series analysis, the GPT-3 language model for processing and analyzing textual data from news and a Kalman filter to improve prediction accuracy by smoothing out the influence of noise on inputs data. The importance of taking into account information from the mass media and social networks to improve the predictive ability of the model is also considered. The purpose of the study is to demonstrate how the integration of various methods and technologies can improve the effectiveness of financial performance forecasting using the example of stock closing prices. The paper presents the results of an experimental evaluation of the proposed approach, showing its advantages over traditional analysis methods. Particular attention is paid to the analysis and discussion of the results of using the proposed model, including the assessment of its effectiveness using both traditional metrics (for example, RMSE) and a specially developed Win Ratio (WR) metric, which allows one to evaluate the model's ability to predict the direction of price changes. The results of the study confirm that the use of a multifactor approach allows one to achieve higher accuracy of forecasts, which can be useful for investors and specialists in the field of financial analysis. The article concludes with a discussion of future research prospects, including the possibility of improving the model through deeper analysis of textual information and the integration of additional data, such as global economic indicators and information verified using blockchain technology. The proposed approach and the results obtained demonstrate significant potential for the development of forecasting methods in financial markets and can contribute to making more informed investment decisions.

**Keywords:** Machine Learning, Sentiment Analysism, LSTM, Kalman Filter, Time Series Analysis, Stock Closing Price Prediction

## Introduction

Investing is an important financial planning tool that helps investors earn profits and ensure financial stability in the future. However, investors face uncertainty and risk when making investment decisions. In this regard, many investors resort to using various investment forecasting methods.

Currently, neural networks are one of the most promising methods for predicting investments. Neural networks are a mathematical model that imitates the functioning of the human brain and is capable of learning based on the data provided. At the moment, there is a lot of research on this topic using various architectures and technologies. Most of them use the same model trained with different data from different markets. Among them there are also those that use a combined approach using several models and technologies in order to take into account more factors, thereby increasing accuracy.

We believe that combining different technologies is the next step not only in forecasting but also in the development of neural networks. Based on the articles we've read, our own research and our practice, we propose an approach that involves using recurrent neural networks (LSTMs) to analyze stock price time series, as well as the

GPT-3 language model to analyze sentiment in news headlines. In addition, we will apply a Kalman filter to filter out noise and improve the accuracy of predictions. In addition to the already ordinary historical data on prices, which is no longer new news, we will also add data from company reports.

The purpose of this article is to study the effectiveness of the proposed approach and its potential for application in real investment conditions. We will present the results of experiments on real stock market data and compare them with traditional analysis methods. Our goal is to demonstrate how the integration of various technologies and analytical approaches can improve the quality of investment decisions and help investors make more informed and informed decisions in the stock market.

## *Literature Review*

Our research began with a literature review, or rather, we took it up after that. In this section, we would like to talk about the articles that we studied. Table 1 provides a summary of the articles reviewed. We think that after reading these articles there should be no questions about why these particular models were chosen.

The most popular architecture for predicting cryptocurrencies is the LSTM model. In the study described in the study Ammer and Aldhyani (2022), he used LSTM, which has been used to predict the value of various cryptocurrencies, such as electro-optical systems, AMP, etc. The results showed that the LSTM algorithm shows excellent results in predicting the prices of all the cryptocurrencies in question, making it one of the most effective tools to perform this task. The LSTM model showed high prediction accuracy for all cryptocurrencies used to predict future closing prices within 180 days. Pearson's correlation measurement was used to estimate the accuracy of predictions that measured the relationship between predicted and actual values in training and testing. The LSTM algorithm showed the highest correlation values in predicting XRP scores: 96.73 in training and 96.09% in testing. These results confirm that the LSTM model can accurately predict the price of a cryptocurrency. The application of this model can significantly improve strategic planning and decision-making in cryptocurrency trading, providing more accurate and reliable forecasts of market prices. Professors at Vilnius Gediminas Technical University in their joint paper (Ma *et al*., 2022) also use this model. For the research, the authors created two portfolios: The IT portfolio and the ETF portfolio. The S&P 500 was used as a benchmark to compare changes in the market and portfolios. These financial instruments were chosen because of the large-capitalization of IT companies and their net assets in ETF indexes. These instruments were chosen because of large changes in the price of stocks or indices. They adopted a prediction model in their study. This study also shows that a neural network or other artificial intelligence algorithms can help determine strategies because the neural network provides a measure of the probability of bad or good situations.

Other researchers (Ma *et al*., 2022; Al-Nefaie and Aldhyani, 2022) have compared the performance of LSTM with other architectures. In Ma *et al*. (2022) article it is compared to DNN deep neural networks. The results showed that over 35 repetitions, the learning error in the MA'ACH model decreased from 8 to almost 0. In 2022, in 20 days, the average real return on Maa securities shares was 13.89. The average return for IA-based predictive models was 14.01, 13.60 for ia-based models and 13.78 for models based on ia. The cognitive repetition model meets the requirements of the repetition learning process, reduces the number of repetitions and saves time on learning.

**Table 1:** Articles based on time series analysis

| No. | Article | Exchange | Period | Time | Methodology | Performance indicators |
|---|---|---|---|---|---|---|
| 1 | Zaini *et al*. (2020) | Malaysia | 2008-2017 | Month | DES by Holt, ANN | MAD, MSE и RMSE |
| 2 | Ammer and Aldhyani (2022) | Cryptocurrency | 2015-2022 | Day | LSTM | MSE, RMSE и NRMSE |
| 3 | Chen *et al*. (2021a) | China | 1990-2021 | Day | CNN-BiLSTM-ECA | MSE, RMSE и MAE |
| 4 | Chen *et al*. (2021b) | Currency | 2021 | Day | Deep learning algorithms and dual-target measurement optimization-based | MSE, RMSE и MAE on NSGA-II |
| 5 | Mazraeh *et al*. (2022) | Tehran | 2011-2020 | Month | RF, SVM, MOGVO и NSGA-II | ROC, SMA, EMA, WMA и MACD |
| 6 | Ma *et al*. (2022) | China | 2022 | Month | LSTM, RNN, BPNN | EMA, WMA и MACD |
| 7 | Miao and Huang (2022) | Gold and cryptocurrency | 2016-2021 | Day | SMA-LSTM | AV |
| 8 | Al-Nefaie and Aldhyani (2022) | Saudi Arabia | 2018-2020 | Day | LSTM, MLP | MSE, RMSE, R-squared и NRMSE |
| 9 | Hu *et al*. (2020) | China | 2007-2019 | Day | ARMA, GARCH | MSE, MAE |
| 10 | Maknickienė and Sabaliauskas (2019) | Mixed | 2014-2019 | Day | LSTM | Точность |
| 11 | Kang *et al*. (2020) | South Korea | 2013-2017 | Day | GA, ANN, regression model | MAPE, RMSE |
| 12 | Agrawal *et al*. (2022) | India | 2008-2018 | Day | EDLM | MSE |

In Chen *et al*. (2021a); Miao and Huang (2022), the authors combined the LSTM architecture with other technical tools. In Miao and Huang (2022), the authors used a simple moving average model to predict the raw data and then only LSTM. In the study, Chen *et al*. (2021b), the authors proposed the CNN-BiLSTM-ECA model, which combines the use of dual LSTM and Convolutional Neural Network (CNN), dual Long-term and Short-Term Memory network (BiLSTM) and Attention Mechanism (AM). In particular, CNN is used to extract deep characteristics from these files, which reduces high noise and nonlinear effects. The bilstm network is then used to predict stock prices based on the deep characteristics obtained. In addition, a new Effective Channel Attenuation (ECA) module will be added to the model, which will increase the sensitivity of the network to critical functions and important data. Finally, extensive experiments were carried out with three Cadastral data sets: The Shanghai Composer index, the Chinese Unicom and CSI 300. Compared to existing methods, the results of the experiment confirmed the effectiveness and feasibility of the proposed CNN-BiLSTM-ECA model, which can become an important decision-making tool for investors.

Other articles use different but equally interesting techniques. For example, the article used the method of Double Exponential Holt Smoothing (DES) and Artificial Neural Network (ANN). The study used the monthly closing price data of Malaysian stock markets such as AM001 Berhad, CI002 Berhad, hl003 Berhad and Pb004 Berhad to predict stock market prices between 2008 and 2017. The results showed that the statistical forecasting method surpasses the ANN Holt model and can accurately predict future price changes in Malaysian stock markets using real values. In this way, the Holt method helps to reduce potential risks and increase profits. However, this study is limited because the two forecasting models discussed here cannot guarantee successful stock price forecasting. In order to obtain sufficient information and make forecasts, investors should use statistical methods, technical analysis and other methods involving artificial intelligence models.

The article Kang *et al*. (2020) presents three models for predicting the value of real estate sales at auction using artificial intelligence methods and statistical methods: Regression model, Artificial Neural Network model (ANN) and Genetic Algorithm (GA). The GA model, grouped by the estimated value of the auction, turned out to be the most profitable of all the tested models. These empirical results show that correct group process criteria play an important role in improving predictive accuracy. The GA model also provides valuable advice for real estate investors and real estate fund managers.

The study Chen *et al*. (2021a) provides two algorithms for predicting currency volatility: A deep learning algorithm and an investment portfolio optimization algorithm based on NSGA-II. Compared to traditional exchange rate forecasting algorithms, the deep learning model provides more accurate results, while the NSGA-II-based model optimizes investment portfolio selection and provides investors with a reasonable investment plan. In general, the model presented in the article will help investors make more informed decisions in the foreign exchange market. The model was implemented through self-programming and gave excellent results. Then you can try to use complex learning algorithms, including autoencoders, deep learning networks and convolutional neural networks.

In the article Mazraeh *et al*. (2022), another researcher concluded that the MOGWO algorithm is superior to NSGA-II. in the analysis, the MOGWO algorithm showed a portfolio return of 133.13 at a risk of 3.346%, while the NSGA-II algorithm showed a return of 107.73 at a risk of 1.459%. A comparison of decision-making methods shows that the MOGWO algorithm is very effective in optimizing the stock portfolio.

The authors of the article Hu *et al*. (2020) use the time series prediction method to correct and predict the daily logarithmic turns of the four main banks. The ARMA and GARCH models are built and compared empirically. The results showed that the GARCH model is better in terms of compatibility effect and the ARMA model is better in terms of predictive effect. Overall, the ARMA model shows good results in volatility and swing prediction, but the GARCH model eliminates conditional heteroscedasticity better.

In the study Agrawal *et al*. (2022) the authors developed an advanced evolutionary research (EDLM) model to identify stock price trends using STDs. The proposed model implements a deep learning model to construct the correlation tensor concept. The Evolutionary approach to Deep Learning (EDLA) does not depend on the market, as it can be applied to securities and stock indices. In this study, potential indicators are determined by the dynamics of the LSTM. Two reference algorithms EDLA and machine learning are used to predict stock market trends through technical indicators. The experiment was carried out on three well-known exchanges in India and the results showed the superiority of the proposed model. In addition, the deep learning model can be improved using various optimization methods and other technical indicators.

The researchers conducted a study in which they used the LSTM architecture in combination with text sentiment analysis and the Kalman filter to predict investments. They used stock prices and news data to train the model. The results of the study showed that using LSTM in combination with text sentiment analysis and the Kalman filter can improve the accuracy of investment forecasting.

**DOI: 10.3844/jcssp.2024.1388.1396**

**Table 2:** Articles based on the analysis of the text and identifying its mood

| No. | Article | Sources | From | Period | Exchange | Methodology | Performance indicators |
|---|---|---|---|---|---|---|---|
| 1 | Mazraeh *et al*. (2022) | cnYES | News | 2016-2017 | Gold | GA-LSSVR, TF-IDF | MAPE |
| 2 | Jiang *et al*. (2021) | East Fortune Stock bar Forum, Python | Investors | 2016-2017 | SSE | SVM, BP | ERS |
| 3 | Cai *et al*. (2020) | Chinese internet | Investors and consumers | 2020 | China's energy market | BERT-BiLSTM | F1, accuracy and completeness |

All of the above articles are based on the behavior of the stock price and its various patterns. But also, one of the important factors of price changes is the mass media, especially in our digital times. There are several articles on this subject (Yuan *et al*., 2020; Jiang *et al*., 2021; Cai *et al*., 2020). Which are aimed at analyzing the text and identifying its sentiment. Table 2 shows a summary of the analysis. For example, in the article Jiang *et al*. (2021) there is a clear internal relationship between investor sentiment and stock indices, which makes it possible to predict general fluctuations in prices on the stock exchange. This study used investor feedback on specialist forums such as the East Fortune Stock Bar forum. Emotional texts were collected from online forums using the Python programming language, which was then used to process them and search for text to obtain financial terminology, which ensured high data accuracy. The correlation analysis method was used to create a comprehensive index of investor sentiment and eliminate bias in forecasting based on sentiment data. Next, a stock index forecasting model was developed that allows for better accuracy in predicting stock market trends using machine learning. The study found that the Vector reference Method (SVM) model of the SSE composite stock index was more effective in predicting the stock index.

The article combined time series analysis, text sentiment analysis and the Kalman filter. An off-the-shelf BERT model was used for text analysis and the LSTM architecture was used for time series analysis and prediction. Thanks to the Kalman filter, the accuracy was improved. The authors note the effectiveness of this approach by attaching it to the results of their research.

This was followed by an article by Li and Pan (2022). In this study, a team model combining different engineering tools was developed. The model combines time series analysis and text analysis. Although no Kalman filter was used, the results of the study were very impressive. The model uses a mixed-team learning method by combining two repeating neural networks and a fully connected neural network. The aim of this study is to demonstrate that in-depth team training methods can more effectively predict future stock price trends and help investors make informed investment decisions about traditional methods. Unlike most existing models, which use a single neural network to predict the cost of effects, the proposed model uses multiple models. This is important because stock prices are highly volatile and depend on many factors, making it difficult to make accurate predictions using a single model. The main contribution of this study is to provide a model that uses several additional models, each of which analyzes different characteristics of the data and reduces noise.

## Materials and Methods

This section will describe the system architecture of my proposed method. It describes the GPT-3 model, LSTM for stock price prediction and Kalman filter, data collection and preprocessing.

### *LSTM Neural Network*

Long Short-Term Memory (LSTM) is a type of recurrent neural network architecture designed to solve the fading gradient problem that occurs when training deep neural networks. LSTMs are particularly effective in processing and predicting sequential data such as speech, text and time series.

The LSTM architecture (Fig. 1) is based on the idea of having "memory cells" that can store information over time. These memory cells have input, forgetting and output gates that regulate the flow of information into and out of the cell.

After reviewing the literature, it was clear that it is currently the best choice for time series forecasting. It is the basis of our method.

### *GPT-3*

Generative Pre-trained Transformer 3 (GPT-3) is a powerful language of artificial intelligence created by OpenAI. It is based on the transformer architecture (Fig. 2) and is designed to generate natural language, understand text and answer questions.

It learns from huge amounts of textual data to learn language patterns and make connections between words and sentences. It can then be used to automatically generate texts, reports, news, articles and even program code.

GPT-3 can also be used for language processing tasks such as translation, text summarization, classification and tone analysis. It can be integrated into a variety of applications and is used in many fields, including business, education, media and science.

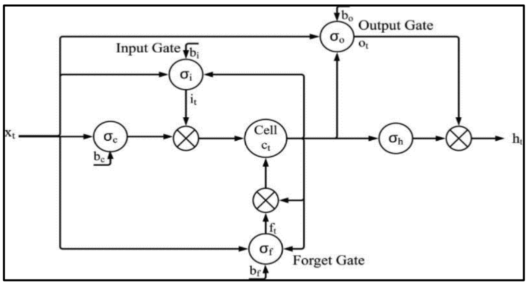


**Fig. 1:** Demonstration of the LSTM model by Gao *et al*. (2017), where *xt* is the input vector at time *t*, *ht* is the output vector, *it* is the union state, it is the input gate vector, *ft* is the forgotten gate vector, *ot* is the output gate vector and *σi σf, σo, σc, σh* are activation functions

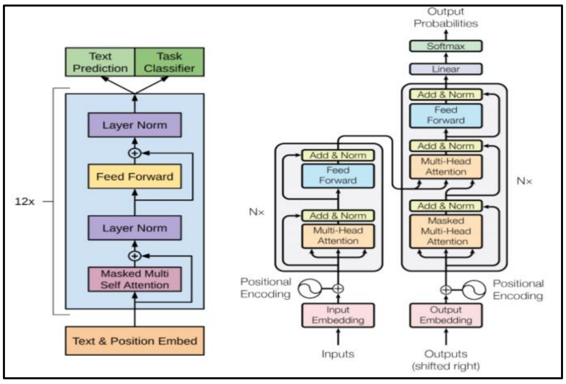


**Fig. 2:** GPT-3 architecture

AI can be used to analyze tone and determine the mood of a text. It can be useful for monitoring public opinion, analyzing user feedback, determining the tone of news articles and many other tasks.

Although there was no research on this model in the literature review, the decision to use it was not made lightly. We tried a variety of models to analyze the sentiment of news headlines. Its ability to analyze text was noticeably better than other models and it could be pre-trained for our tasks, that is, for news headlines for specific companies.

### *Kalman Filter*

The Kalman filter is an optimal linear filtering algorithm that is used to estimate the state of a dynamic system based on noise observations. It was developed in the 1960s by Rudolf Kalman and Steven Booker and has since found wide application in fields such as aeronautical and space engineering, robotics, financial applications, etc.

The principle of the Kalman filter is to use information about the previous state of the system and the current observations to obtain an optimal estimate of the current state of the system. The algorithm (Fig. 3) uses mathematical models to describe the system dynamics and statistical methods to model the noise in the observations. This yields more accurate estimates of the system state than if only the current observations were used.

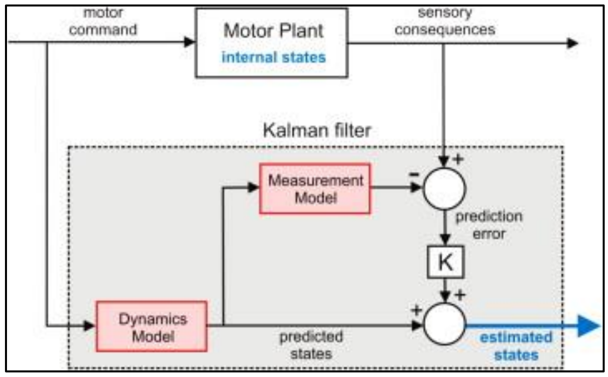


**Fig. 3:** Kalman filter architecture

Since we work with real data, the presence of noise is inevitable. There are many ways to reduce noise, but the Kalman filter has proven to be the best.

### *Data Collection*

For a good model, the data must be clear and truthful. It is also important not to "clutter" the model with unnecessary data. There are many factors that can influence a company's share price, but the most important thing is to understand exactly how pricing works. The price is primarily influenced by "buyers" and their trust in the company. Therefore, any high-profile news headlines, comments from media and authoritative people greatly influence people's opinions about this company and, accordingly, the price. Therefore, it is important for a company to maintain its "image" if they want to develop steadily. To do this, companies periodically provide reports on the current state of affairs of the company in order to show investors that they are doing well. Of course, there are many ways a company can cheat to improve its performance on paper. In the short term, these machinations help support the price for inexperienced investors, but in the long term this does not work and a price decline follows. Be that as it may, any information that is provided specifically or not affects the opinion of “buyers” and therefore the price. We believe that this data is correlated, so we will use it for forecasting.

According to our proposed method, three categories of data are required. This is historical information about stock transactions, in particular closing prices, news and quarterly reports. The algorithm for collecting and processing data is presented in Fig. (4), more details will follow. It was decided to use data on AAPL shares for the dataset. Historical data was taken from Kaggle, quarterly reports were taken from Apple's official website and news headlines were obtained using NewsApi. All data covers the period from January 1, 2015, to December 31, 2022.

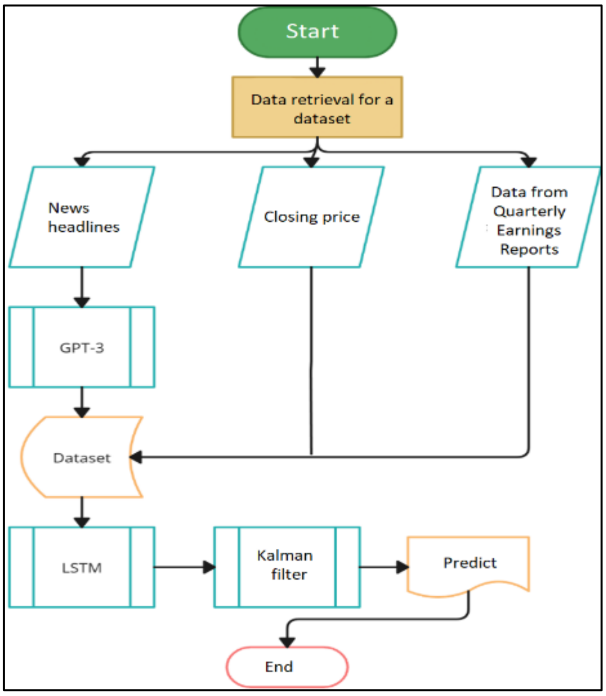


**Fig. 4:** Algorithm of the model

### *Data Pre-Processing*

To begin with, using GPT-3, news headlines were analyzed and classified into positive, negative and neutral. Then, using the label encoder method, this data was converted into numeric values, where 1 is positive, 0.5 is neutral and 0 is negative. For each day, the average value was recorded. On days when there was no news, it was filled with a value of 0.5. From information on stock transactions, only closing prices, which are the target parameter, are taken. Two columns were taken as fundamental data: Net profit and balance at the end of the period. Since the stock exchange does not work on weekends and news may be released on weekends, there were a lot of empty values. Also, reporting is provided once in a certain period of time, which means that there will be many empty days there too. In order to convert all this data into one plane, it was pre-filled. Weekend closing prices were filled with the value from Friday. In the case of Apple, the company provides reports quarterly, which means that once every three months we will have data and have to fill it out. The report affects the price of subsequent days. Therefore, the empty fields were filled not in the logic of which days this refers to, but in the logic of influence. That is, empty fields were filled with the nearest previous value. Using the Kalman filter the noise from the data was reduced. Next, using MinMaxScaler, closing prices and data from reports were scaled to the range from 0-1.

### *Training the LSTM Model*

The dataset was divided into training and testing material, 80-20% respectively. Using the first one, an LSTM model was trained. Using the test set, the accuracy of the forecast will be assessed. The model will consist of three LSTM layers that will be connected in series with each other and then connected to a fully connected layer.

Once the model is trained, you can use it to predict the closing price for the next week. To do this, you need to take the last 30 days of data and feed it into the model's input to get a forecast of the closing price. The accuracy will be assessed by the RMSE metric, a script was also written that will calculate the Win Ratio (WR) which will show how often the model guesses the direction of movement (whether the price will rise or fall). We believe that WR should show how well the model will guess. After all, due to the use of different scaling, any model can show good indicators that will not work in real cases and the WR indicator should correct this.

## Results

Since the purpose of this study is to demonstrate that combining different models to cover large factors improves forecasting accuracy, the separate use of an LSTM model, a model with Kalman filter integration and a model incorporating news analysis was evaluated. Next, we presented the results of a comprehensive model combining all three approaches.

Our analysis showed that while standard metrics such as RMSE and MSE are widely used to evaluate machine learning models, their applicability may be limited due to differences in data scaling. As an alternative, we proposed using the Win Ratio (WR) metric, which, in our opinion, more accurately reflects the effectiveness of the model in real conditions. Despite this, for completeness, we have also presented the RMSE results.

The final model demonstrated a predictive ability with an accuracy of 82.7% as shown in Table 3, indicating its high performance. Also in Fig. (5) you can see that the model predicts the price movement trend well. However, there are ways to further improve the model.

As a potential direction for improvement, a modification of the model to predict the direction of trends by adapting the dataset is proposed. By aggregating the monthly averages and adding a new target, the growth rate (where 1 represents an increase in price compared to the previous month and 0 indicates no increase), a classification model can be trained to predict the change in the average closing price in the next month. This approach can significantly increase the practical value and accuracy of predictions.

**Table 3:** Model evaluation

| Metrics | LSTM | LSTM + KF | LSTM + GPT-3 | Final model |
|---|---|---|---|---|
| RMSE | 4.56 | 4.42 | 4.48 | 4.33 |
| WR | 0,682.00 | 0,689.00 | 0,755.00 | 0,827.00 |


**Fig. 5:** Results

## Discussion

The model is already showing its efficiency and usefulness for long-term investment. But there are a number of directions for further development and improvement of the forecasting model. One promising area is a deeper analysis of media and social network data. The importance of this aspect is due to the ability of public opinion and sentiment to influence the dynamics of financial markets:

1. Extended analysis of the opinions of investors and users of social networks. Including comments from well-known investors and ordinary users on social networks under financial accounts' posts in the analysis can provide additional forecasting signals. Analyzing such data requires the development of sophisticated algorithms to identify and classify opinions, which can significantly improve the accuracy and relevance of forecasts
2. More detailed classification of emotions in media content. The current approach to sentiment analysis, which classifies content as positive, negative, or neutral, can be greatly improved by dividing emotions into more specific categories. This will allow us to take into account more subtle nuances in the perception of information by different groups of investors and consumers. Developing methods to automatically detect emotions such as optimism, pessimism, excitement and confidence could lead to significant improvements in the quality of forecasts
3. Integration of global economic indicators. An additional area of improvement for the model is the inclusion of global economic indicators such as consumer confidence indices, unemployment data, interest rates and other macroeconomic indicators. This will help the model take into account the overall health of the economy and the potential impact of external factors on the market
4. Application of blockchain technology for data verification. In an era of information overload and data manipulation, source verification becomes critical. Using blockchain technology to verify the authenticity of social media data can provide a more reliable basis for sentiment and opinion analysis
5. The implementation of these areas will require an integrated approach, including both improving machine learning algorithms and developing new methods for collecting and analyzing data. However, the potential increase in forecast accuracy and relevance makes this effort worthwhile. It is expected that the integration of these approaches will significantly expand the ability to forecast trends in financial markets, increase their accuracy and, as a result, improve the quality of decisions made in the market

## Conclusion

In this study, an integrated approach to forecasting stock closing prices based on the use of recurrent neural networks (LSTM), sentiment analysis using the GPT-3 model and the Kalman filter was presented and evaluated. The pilot study confirmed the hypothesis that the integration of diverse data sources and analysis methodologies can significantly improve the accuracy of financial performance forecasts.

The main contribution of this study is the development and testing of a multifactor forecasting model, which includes not only traditional financial indicators but also a comprehensive analysis of sentiment identified on the basis of news publications. The effectiveness of the proposed model was demonstrated and recommendations were provided for further directions of research, including in-depth analysis of the emotional tone of mass media messages and social commentary, as well as the integration of global economic indicators.

In addition, the relevance of developing new methods of data verification and authentication using blockchain technology was discussed, which could open up new opportunities for increasing the reliability and transparency of the data used.

The results of the study highlight the importance of an integrated approach to the analysis of financial markets and can serve as the basis for the development of more advanced forecasting systems adapted to the conditions of high volatility and uncertainty in modern financial markets. Future research in this area could focus on experimental validation and optimization of the proposed methods under different market conditions and for different asset categories.

Thus, the approach proposed in the work and the results obtained may be of interest to the scientific community involved in forecasting in financial markets, as well as to practitioners in the field of investment and asset management seeking to improve the efficiency of investment strategies based on analytical models.

## Acknowledgment

We would like to thank Al-Farabi Kazakh National University for the conditions and opportunities for research provided.

## Funding Information

The authors did not receive any monetary assistance for the investigation, composition, and/or publication of this manuscript.

## Author's Contributions

**Assel Abdildayeva:** Provided supervision and guidance throughout the semantic research component of the project.

**Galymzhan Nurtugan:** Contributed to the conceptualization, formal analysis, methodology, coding, validation, and initial drafting of the study.

**Guldana Taganova:** Contributed to the testing and analysis of results.

**Ardak Akhmetova:** Contributed to annotation and interpretation; performed revisions and corrections to the manuscript prior to submission.

## Ethics

This study accurately and comprehensively represents the authors' research and analysis. It acknowledges the valuable contributions of co-authors and co-researchers. The findings are appropriately contextualized within the existing body of research. All references and related works are properly cited.

### *Conflict of Interest*

The authors hereby declare their complete independence from any organization or entity that may have a financial or non-financial interest in the subject matter or materials discussed in this manuscript.

## References

Agrawal, M., Shukla, P. K., Nair, R., Nayyar, A., & Masud, M. (2022). Stock prediction based on technical indicators using deep learning model. *Computers, Materials & Continua*, *70*(1). https://doi.org/10.10.32604/cmc.2022.014637

Al-Nefaie, A. H., & Aldhyani, T. H. (2022). Predicting close price in emerging Saudi Stock Exchange: Time series models. *Electronics*, *11*(21), 3443. https://doi.org/10.3390/electronics11213443

Ammer, M. A., & Aldhyani, T. H. (2022). Deep learning algorithm to predict cryptocurrency fluctuation prices: Increasing investment awareness. *Electronics*, *11*(15), 2349. https://doi.org/10.3390/electronics11152349

Cai, R., Qin, B., Chen, Y., Zhang, L., Yang, R., Chen, S., & Wang, W. (2020). Sentiment analysis about investors and consumers in the energy market based on BERT-BiLSTM. *IEEE Access*, *8*, 171408-171415. https://doi.org/10.1109/ACCESS.2020.3024750

Chen, J., Zhao, C., Liu, K., Liang, J., Wu, H., & Xu, S. (2021a). Exchange Rate Forecasting Based on Deep Learning and NSGA-II Models. *Computational Intelligence and Neuroscience*, *2021*(1), 2993870. https://doi.org/10.1155/2021/2993870

Chen, Y., Fang, R., Liang, T., Sha, Z., Li, S., Yi, Y., ... & Song, H. (2021b). Stock Price Forecast Based on CNN-BiLSTM-ECA Model. *Scientific Programming*, *2021*(1), 2446543. https://doi.org/10.1155/2021/2446543

Gao, T., Chai, Y., & Liu, Y. (2017, November). Applying long short-term momory neural networks for predicting stock closing price. In *2017 8th IEEE International Conference On Software Engineering And Service Science (ICSESS)* (pp. 575-578). IEEE. https://doi.org/10.1109/ICSESS.2017.8342981

Hu, Y., Tao, Z., Xing, D., Pan, Z., Zhao, J., & Chen, X. (2020, August). Research on stock returns forecast of the four major banks based on ARMA and GARCH model. In *Journal of Physics: Conference Series* (Vol. *1616*, No. 1, p. 012075). IOP Publishing. https://doi.org/10.1088/1742-6596/1616/1/012075

Jiang, Y., Zuo, Y., & Yang, Y. (2021, July). Machine learning and artificial intelligence applied in the research of the emotions impact on forecasting. In *Journal of Physics: Conference Series* (Vol. *1982*, No. 1, p. 012052). IOP Publishing. https://doi.org/10.1088/1742-6596/1982/1/012052

Kang, J., Lee, H. J., Jeong, S. H., Lee, H. S., & Oh, K. J. (2020). Developing a forecasting model for real estate auction prices using artificial intelligence. *Sustainability*, *12*(7), 2899. https://doi.org/10.3390/su12072899

Li, Y., & Pan, Y. (2022). A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics*, *13*(2), 139-149. https://doi.org/10.1007/s41060-021-00279-9

Ma, G., Chen, P., Liu, Z., & Liu, J. (2022). The prediction of enterprise stock change trend by deep neural network model. *Computational Intelligence and Neuroscience*, *2022*(1), 9193055. https://doi.org/10.1155/2022/9193055

Maknickienė, N., & Sabaliauskas, D. (2019, May). Investment portfolio analysis by using neural networks. In *International Scientific Conference, Contemporary Issues in Business, Management and Economics Engineering"*. ISBN-10: 978-609-476-161-4.

Mazraeh, N. B., Daneshvar, A., Madanchi zaj, M., & Roodposhti, F. R. (2022). Stock portfolio optimization using a combined approach of multi objective grey wolf optimizer and machine learning preselection methods. *Computational Intelligence and Neuroscience*, *2022*(1), 5974842. https://doi.org/10.1155/2022/5974842

Miao, Z., & Huang, W. (2022). An optimal portfolio method based on real time prediction of gold and bitcoin prices. *Systems Science & Control Engineering*, *10*(1), 653-661. https://doi.org/10.1080/21642583.2022.2096149

Yuan, F. C., Lee, C. H., & Chiu, C. (2020). Using market sentiment analysis and genetic algorithm-based least squares support vector regression to predict gold prices. *International Journal of Computational Intelligence Systems*, *13*(1), 234-246. https://doi.org/10.2991/ijcis.d.200214.002

Zaini, B. J., Mansor, R., Md Yusof, Z., Gabda, D., & Wong, K. S. (2020). Comparison of double exponential smoothing for holt's method and artificial neural network in forecasting the Malaysian banking stock markets. *ASM Science Journal*, *13*, 1-5. https://doi.org/10.32802/asmscj.2020.sm26(1.4)