JOURNAL OF BASIC SCIENCE AND ENGINEERING

ACCURATE STOCK MARKET HIERARCHICAL TREND PREDICTION FORECASTING WITH ARTIFICIAL OPTIMIZATION ALGORITHM

[1]Vanama Manisha [2]Dr. Ratna Raju Mukiri [3]S. Amarnath Babu
[1]M. Tech Scholar, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
[2]Associate Professor, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
[3]Associate Professor, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
e-mail: mukiriratnaraju001@gmail.com

ABSTRACT: The landscape of stock market prediction is undergoing a profound transformation driven by technological advancements and data-driven methodologies. Within this shifting paradigm, artificial intelligence (AI), particularly deep learning (DL) is emerging as a transformative tool to enhance predictive accuracy stock market trend, a textual embedding construction method is proposed to encode multiple textual features, In recent years machine learning and deep learning have become popular methods for financial data analysis, including financial textual data, numerical data, and graphical data. An optimized deep LSTM network with the ARO model is created to predict stock prices. DJIA index stocks are used as the dataset. LSTM-ARO is compared with one artificial neural network (ANN) model, three different LSTM models, and LSTM optimized by Genetic Algorithm (GA) model. New hybrid model, termed Hierarchical Decomposition-based Forecasting Model (HDFM), to decompose and forecast stock prices in a hierarchical fashion. The first combined sub-series is subjected to a second decomposition with variational mode decomposition (VMD). To enhance sentiment analysis accuracy, this study proposes text classification using Bidirectional Encoder Representations from Transformers (BERT) and its variants for natural language processing. Experimental results demonstrate the effectiveness of combining BERT with Convolutional Neural Networks (CNN). The purposedesign philosophy ensemble deep learning technologies future stock price trends more effectively and can better assist investors in making the right investment decision than other traditional methods.

Index Terms: Stock Market Prediction, Deep Learning (DL), Sentiment Analysis, Fake News Detection, Artificial Intelligence (AI), Ensemble learning · Statistical finance,

INTRODUCTION

India has emerged as one of the world's fastest-expanding markets, with two leading stock exchanges: the Bombay Stock Exchange (BSE) & the National Stock Exchange (NSE). The NSE is India's largest stock exchange, having a market capitalization of more than $3.7Trillion USD as of March 2023 [1]. Stock market prediction is an interesting and valuable research task for both industry and academia [2].The development of big data, AI techniques and machine learning methods advanced methods using machine learning even deep learning models have been widely applied to stock market prediction tasks recently [3]. People usually have a short memory about stock factors. Hence, determining a suitable historical window size is important to correctly

predict stock prices [4]. In recent years artificial intelligence (AI) has become more prevalent in the financial industry including the stock market. AI algorithms have the capability to analyze vast amounts of data and make predictions based on that analysis. This can be useful for predicting stock prices, identifying trends, and making investment decisions [5]. These include autoregression (AR), moving average (MA), autoregressive moving average (ARMA) [1], autoregressive integrated moving average (ARIMA) [2], generalized autoregressive conditional heteroskedasticity (GARCH) [3], and linear regression [4]. Although these methods have achieved reasonable predictive performance the assumption of linearity limits their predictive ability, especially for stock prices with high volatility [6]. The emergence of Machine Learning and its powerful algorithms have brought new advancements to stock market research and prediction by enabling the analysis of vast amounts of market data [7]. In stock market prediction the target can be the future stock price, price volatility market trends. Prediction methods can be categorized into two types: dummy prediction and real-time prediction [8].

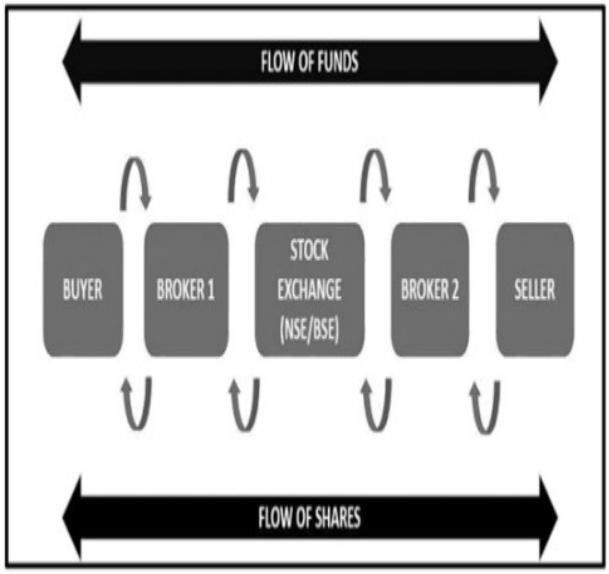

Figure1. Flow of stocks and shares in a common stock market

1. RELATED WORK

The Sayavong Lounnapha proposes a prediction model based on Convolutional Neural Networks (CNNs) applied to the Thai stock market [9]. The model demonstrates exceptional capability in learning the market behavior and accurately predicting stock price trends reports elevated prediction accuracy, highlighting the potential of CNNs in the field of finance [10].In recent years, this field's research landscape has witnessed the ascendancy of deep learning models. Concurrently data science, datasets, visual-based methods, and associated elements persist in their pivotal roles. Consequently, a systematic literature review (SLR) is imperative to advance scientific comprehension and knowledge within this domain [11]. a stock prediction model was developed based on the sentiments of the company's specific topics. The proposed approach as well as existing topic models are used to automatically extract topics and associated

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

sentiments from a message board. In a different approach [12]. One of the significant challenges of using artificial intelligence to predict the stock trend with news is the weight of news events in it and the extent to which each news event affects the trend of stock. Hu [6] proposed using a hybrid attention network (HAN) [13]. ARO is a meta heuristic algorithm that is inspired by the behavior of the rabbits in the nature. It is used to find approximate solutions to optimization and search problems. The basic idea is to use techniques inspired by survival strategies of rabbits, such as foraging, hiding, to generate new solutions to a problem [14]. To alleviate the challenge of selecting hyper-parameters the proposed adaptive SVR adopted a dynamic mechanism and a particle swarm optimization algorithm to adjust the parameters. The experimental results showed the efficacy of the adaptive SVR [15] employed a support vector machine (SVM) in combination

2. SYSTEM ANALYSIS

Sentiment analysis is a pivotal aspect of stock market prediction, bridging the gap between investors and market behavior and Natural Language Processing (NLP) and machine learning algorithms this analytical approach dissects financial news and social media data to classify sentiments as negative, neutral, or positive [18]. The proposed cooperative architecture suggested LSTM-News and GRU-News models as equally effective. Limitations encompassed context-specific findings lack of external validation, and reliance on sentiment analysis the vital role of DL models in stock market forecasting, focusing on risk management profitability metrics predictor techniques and trading strategies [19]. A popular research area in Natural Language Understanding (NLU) is social media sentiment analysis, which identifies and categorizes opinions expressed in news, articles, tweets, blog posts, or any other texts [27], [28]. In the field of stock market prediction, many researchers enhance stock movement [20]. GA is a computational method that mimics the process of natural selection to find solutions to optimization and search problems. The idea behind it is to use a population of candidate solutions, called individuals, and repeatedly apply genetic operators selection crossover and mutation to generate new individuals.

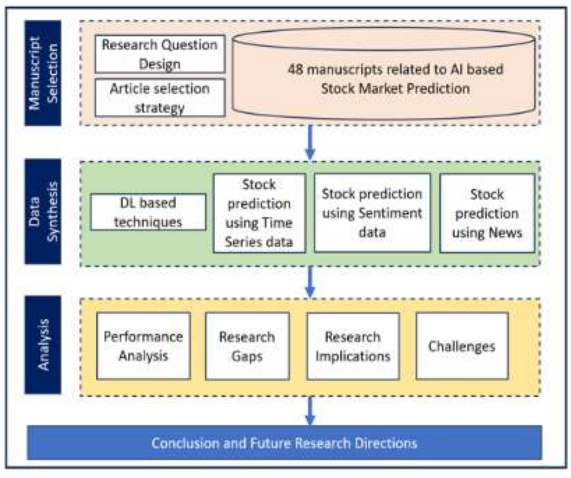

Figure 2. Manuscript workflow

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

3. PROPOSED SYSTEM

Our proposed method aims to leverage real stock market time series data and stock news texts data to predict the stock price trend. We introduce a novel stock market trend forecasting method to investigate and compare the performance of multiple textual features from social media news texts analysis [21]. The pre-processing, all the null data are removed from the dataset and all news data and stock data are combined market only opens and closes during the weekdays therefore weekends are not included in the dataset [22]. Our method consists of three stages, each with multiple steps.

Stage 1: We decide on the optimal number of topics for stock news texts and generate the topic features MT T to represent the topic distribution for each news item Additionally a piece of text document can also be represented by its topic distribution using the Latent Dirichlet Allocation (LDA) framework.

Stage 2: We construct the news textual embedding MT consisting of topic features sentiment features we construct financial features stock price data we choose word embedding input features to capture the semantic meaning of the textual data optimal size of sliding window for the financial features

Stage 3: To predict the final stock trend the stock price features XT and news textual embedding MT as the input by the simple but useful way of direct concatenation We compare the performance of different deep learning models, including FNN, CNN and LSTM, through the comparative experiments our proposed method integrating both the stock price features and the news textual embedding features MT constructed [23].

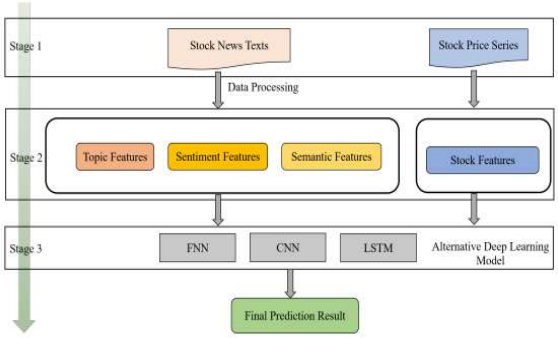

Figure.3. The proposed framework

5. METHODOLOGY

Stock price prediction is classified in the time-series category due to its unique characteristic which means stock price prediction is a continual prediction following the direction of time. The most common techniques used for stock forecasting are statistics algorithms and economics models [24]. We investigate the notions of objectivity and subjectivity features are derived: the average subjectivity and objectivity scores of news articles for each day. Notably, these two scores tend to complement each other with higher subjectivity corresponding to lower objectivity and vice versa.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

In the final preprocessing step, we perform opinion finding mood mining) by assigning sentiment polarity values to tokens in the headline text.

Historical stock market data is collected and preprocessed for analysis.

• The dataset is divided into training and testing subsets for model evaluation.

• Machine learning techniques, including Linear Regression, Moving Average, K-Nearest Neighbors, Auto ARIMA, Prophet, and LSTM, are implemented.

• The models are trained on the training data, and hyperparameters are adjusted as necessary.

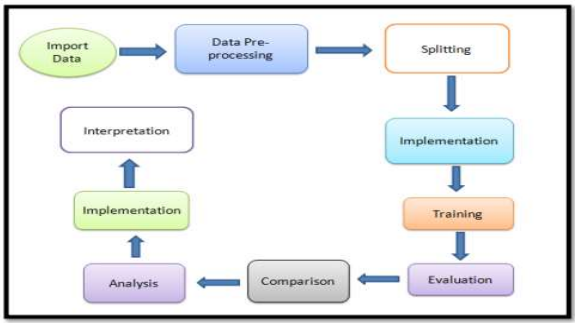

Figure 4. Methodology

CEEMDAN ALGORITHM

The proposed ensemble empirical mode decomposition (EEMD) to deal with non-linear and non-stationary signals This overcomes the problem of EEMD in which the reconstructed signal is incomplete and erroneous. The K-means clustering algorithm [36] is adopted. In addition, to quantify the complexity of different IMFs, the sample entropy [37] is selected as the similarity measurement of two IMFs for K-means.

K-Means Clustering Algorithm

S = {s1,s2, · · · ,sN }, K

S is the set of samples

K is the number of clusters

Input:

Output: Cluster division: C = {c1, c2, · · · ,cK }

1: Initialization: C ← S ELECTR ANDOMS EEDS(S,K)

2: repeat

3: for n = 1, 2, · · · ,N do

4: for k = 1, 2, · · · ,K do

5: if k == arg min k ‖sn − ck‖2 then

6: rnk = 1

7: else rnk = 0

8: end if

9: end for

10: end for

11: for each cluster ck do

12: update cluster centroids as the mean of each cluster
13: end for
14: until all current cluster centroids do not change

Analysis-based stock market prediction

Sentiment analysis is a pivotal aspect of stock market prediction, bridging the gap between investors and market behavior. By harnessing the power of Natural Language Processing (NLP) and machine learning algorithms this analytical approach dissects financial news and social media data to classify sentiments as negative neutral, or positive. Method considers both textual sources as valuable external data for prediction. Accuracy and spam detection are key requirements, especially in the age of spammers utilizing multiple social media accounts for promotional purposes. The removal of spam messages from the dataset is a vital step in ensuring the reliability of sentiment-based predictions in the dynamic world of stock markets.

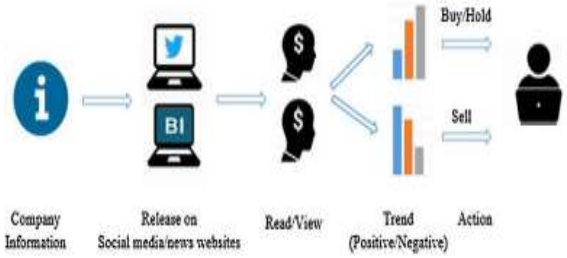

Figure 5: A general plot that illustrates how financial news and social media affect stock market trends.

6. EXPERIMENTAL RESULTS

Experimental results during the experiments, we use the predicted stock price to compare with the actual stock price and calculate the MSE and MPA values. We then use the predicted values to calculate the price fluctuation of the stock on the forecast day; if the predicted stock price increases the output and if the predicted stock price decreases These results will be used for the confusion matrix to calculate various measurements. To further evaluate the blending ensemble model we also deploy a GRU model that is very similar to the sub-model building as a test model. In addition we also recorded the averaging ensemble model and weighted average ensemble model prediction results to contrast with the blending ensemble model.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

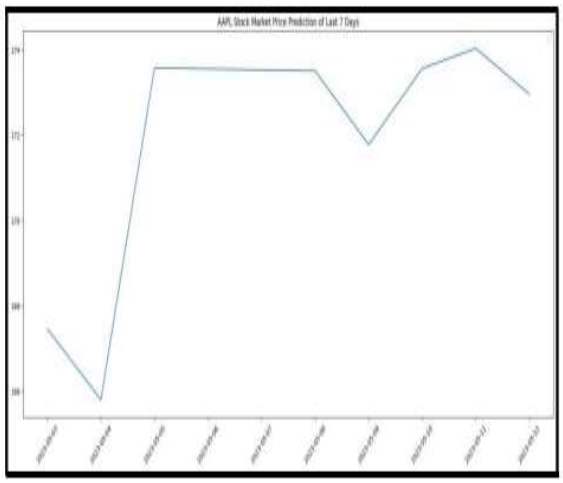

Figure 6. Analysis and prediction of Stock Data

7. CONCLUSION AND FUTURE WORKS

This systematic literature review (SLR) delves into the deployment of deep learning methods for Recent Advances on Stock Markets Predictions. Shedding light on the promising outcomes achieved deep learning and machine learning models in effectively identifying and categorizing NSE and BSE data sentimental analysis and the relationship with fake news. Incorporating both quantitative and qualitative indicators in stock prediction models can provide a more comprehensive understanding of the factors influencing.It may therefore lack the diversity in opinion expressions as compared to gathering news from multiple news sources. Also, there have been many techniques presented in the fields of information fusion and multi-modal pattern recognition. It is difficult for a single model to predict future price trends accurately because a single model often cannot learn all characteristics of the data. We decompose the stock price time series into several sub-series using the CEEMDAN method. To reduce the computation time, sub-series with similar sample entropy are merged using K-means. Future improvements could involve manual annotation and word vector techniques for data quality and exploring alternative sentiment aggregation methods. This system can be integrated into a dynamic trading system or an automated stock market system. It is aimed to use the successful model of this paper in the dynamic trading system. We may use many different ways to improve our current work. Below are possible future research directions.

8. REFERENCES

[1] M. M. Rounaghi and F. N. Zadeh, ‘‘Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: Monthly and yearly forecasting of time series stock returns using ARMA model,’

[2] Dr. Ratna Raju Mukiri. (2018). Estimate Requirement of Online Report Analysis using Random Forest Regression,Vol. 10, 04-Special Issue,.

[3]Allcott H and Gentzkow M (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives 31(2): 211–236
.[4]Almalis, I., Kouloumpris, E., &Vlahavas, I. (2022). Sector-level sentiment analysis with deep learning.Knowledge-Based Systems,258, 109954.
[5]Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions.Journal of big Data,8, 1-74.
[6]Amer, E., Kwak, K. S., & El-Sappagh, S. (2022). Context-based fake news detection model relying on deep learning models.Electronics,11(8), 1255
.[7]Baek, Y., & Kim, H. Y. (2018). ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module.Expert Systems with Applications,113, 457-480
.[8]Bangyal, W. H., Qasim, R., Rehman, N. U., Ahmad, Z., Dar, H., Rukhsar, L., ... & Ahmad, J. (2021). Detection of fake news text classification on COVID-19 using deep learning approaches.Computational and mathematical methods in medicine,2021, 1-14.
[9]Barra, S., Carta, S. M., Corriga, A., Podda, A. S., &Recupero, D. R. (2020). Deep learning and time series-to-image encoding for financial forecasting.IEEE/CAA Journal of Automatica Sinica,7(3), 683-692
[10] A. M. Rather, A. Agarwal, and V. Sastry, “Recurrent neural network and a hybrid model for prediction of stock returns,” Expert Systems with Applications, vol. 42, no. 6, pp. 3234–3241, 2015.
[11] H. T. Thanh and P. Meesad, “Stock market trend prediction based on text mining of corporate web and time series data,” Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 18, no. 1, pp. 22–31, 2014.
[12] A. K. Nassirtoussi, S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo, “Text mining for market prediction: A systematic review,” Expert Systems with Applications, vol. 41, no. 16, pp. 7653–7670, 2014.
[13] S. E. Seker, M. Cihan, A.-N. Khaled, N. Ozalp, and A. Ugur, “Time series analysis on stock market for text mining correlation of economy news,” International Journal of Social Sciences and Humanity Studies, vol. 6, no. 1, pp. 69–91, 2013.
[14] B. Li, K. C. Chan, C. Ou, and S. Ruifeng, “Discovering public sentiment in social media for predicting stock movement of publicly listed companies,” Information Systems, vol. 69, pp. 81–92, 2017.
[15] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, “Stock market prediction using machine learning classifiers and social media, news,” Journal of Ambient Intelligence and Humanized Computing, pp. 1–24, 2020.
[16] M. Javed Awan, M. S. Mohd Rahim, H. Nobanee, A. Munawar, A. Yasin, and A. M. Zain, “Social media and stock market prediction: a big data approach,” MJ Awan, M. Shafry, H.

Nobanee, A. Munawar, A. Yasin et al.," Social media and stock market prediction: a big data approach," Computers, Materials & Continua, vol. 67, no. 2, pp. 2569–2583, 2021.
[17] P. Mehta, S. Pandya, and K. Kotecha, "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning," PeerJ Computer Science, vol. 7, p. e476, 2021
[18] R. Ren, D. D. Wu, and T. Liu, ''Forecasting stock market movement direction using sentiment analysis and support vector machine,'' IEEE Syst. J., vol. 13, no. 1, pp. 760–770, Mar. 2019.
[19] M. Liu, K. Luo, J. Zhang, and S. Chen, ''A stock selection algorithm hybridizing grey wolf optimizer and support vector regression,'' Exp. Syst. Appl., vol. 179, Oct. 2021, Art. no. 115078.
[20] D. Zhang and S. Lou, ''The application research of neural network and BP algorithm in stock price pattern classification and prediction,'' Future Gener. Comput. Syst., vol. 115, pp. 872–879, Feb. 2021.
[21] H.-L. Yang and H.-C. Lin, ''An integrated model combined ARIMA, EMD with SVR for stock indices forecasting,'' Int. J. Artif. Intell. Tools, vol. 25, no. 2, Apr. 2016, Art. no. 1650005.
[22] I. Lauriola, A. Lavelli, and F. Aiolli, ''An introduction to deep learning in natural language processing: Models, techniques, and tools,'' Neurocomputing, vol. 470, pp. 443–456, Jan. 2022.
[23] P. L. T. Mbouembe, G. Liu, J. Sikati, S. C. Kim, and J. H. Kim, ''An efficient tomato-detection method based on improved YOLOv4-tiny model in complex environment,'' Frontiers Plant Sci., vol. 14, Apr. 2023, Art. no. 1150958.
[24] A. Kanwal, M. F. Lau, S. P. H. Ng, K. Y. Sim, and S. Chandrasekaran, ''BiCuDNNLSTM-1dCNN—A hybrid deep learning-based predictive model for stock price prediction,'' Exp. Syst. Appl., vol. 202, Sep. 2022, Art. no. 117123.
[25] B. Gülmez, ''Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm,'' Exp. Syst. Appl., vol. 227, Oct. 2023, Art. no. 120346.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930