Indonesian Journal of Electrical Engineering and Computer Science
Vol. 31, No. 2, August 2023, pp. 1041~1049
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v31.i2.pp1041-1049

Associating deep learning and the news headlines sentiment for Bursa stock price prediction

Zuleaizal Sidek[1], Sharifah Sakinah Syed Ahmad[1], Noor Hasimah Ibrahim Teo[2]
[1]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia
[2]School of Computing, College of Computing, Informatics and Media, Universiti Teknologi MARA (Melaka Branch), Melaka, Malaysia

Article Info

Article history:

Received Dec 8, 2022
Revised Apr 8, 2023
Accepted Apr 16, 2023

Keywords:

Deep learning
Long short-term memory
News sentiment
Sentiment analysis
Stock prediction

ABSTRACT

Accurate stock price prediction is appealing to academics, economists, and financial analysts for its potential to increase profits. Although remarkable progress has been made in stock prediction accuracy, studies to explore the relationship between public sentiments and the prediction of stock price movement based on online news portals in Malaysia context are limited. Therefore, this study aims to determine whether news sentiments influence the movement of the Bursa stock price. The stock prediction model was implemented using long short-term memory (LSTM), with stock data from Bursa Malaysia between January 2017 and April 2022, and the root mean squared error (RMSE) value was calculated. In addition, LSTM prediction model was compared to the decision tree algorithm, and LSTM performed significantly better than the decision tree, particularly when using the New York stock exchange (NYSE) dataset. Furthermore, sentiment analysis was carried out using a Malaysian online news portal's business and financial news. The findings showed that i) news has a significant impact on Malaysian stock market price movement; ii) the RMSE of the LSTM model was improved by adding a parameter (news polarity values); and iii) the RMSE value generated is less than one for every company stock and is influenced by stock price and price change magnitude.

Corresponding Author:

Zuleaizal Sidek
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM)
76100, Durian Tunggal, Melaka, Malaysia
Email: p032220004@student.utem.edu.my

1. INTRODUCTION

A large and complex securities trading environment will cause new investors and traders difficulties in understanding the structure, rules and regulations, daily trading transactions, and selecting potential counters. Stock prediction is a difficult task dealing with global economic uncertainties, government policies and stability, oil prices and fluctuation of stock markets. These factors make stock markets more volatile and create a non-linear trading environment [1]. Therefore, involvement in stock trading is risky and not a place to make mistakes. Traders do not know which stocks to buy and when to sell to maximize their income from stock trading because every single mistake causes a loss of money. Business and economic news related to a particular company's core business often influence the fluctuations of the company’s stock on the market. Any stock prediction application system that ignores the business and economic news often produces inaccurate buying calls.

The prediction of stock prices has been a challenging task for investors and researchers due to the complexity and unpredictability of the stock market. Traditional machine learning models have been widely used in stock price prediction, but their performance is limited due to the inability to capture the complex patterns and relationships in the data. Deep learning, on the other hand, has shown great potential in stock price prediction due to its ability to extract features automatically from the data. In recent years, several studies have explored the use of deep learning in combination with the news headlines sentiment to predict the stock prices of different markets. There is a long-term discussion about stock prediction. Following efficient market hypotheses (EMH), stocks always trade at their fair market value on stock exchanges, preventing investors and traders from profiting from being undervalued or losing money on overvalued stock. Furthermore, stock price movement reflects all information available to investors and traders. The majority (66%) of research has shown that stock prices rely on historical prices and 23% of the research mentioned stock prices rely on fundamental analysis [2]. Therefore, a combination of historical prices, fundamental analysis and textual data (news headlines) might produce a significant stock prediction as suggested by [3] prediction accuracy is increased by incorporating blog sentiment into support vector machine (SVM) models. A review of several techniques used for prediction has been reported in [4].

Predicting stock price movement accurately remains unresolved from the data science perspective. Over the last two decades, machine learning approaches have been explored and investigated to develop a successful prediction model. Data scientists and economists have been interested in this classic yet complex topic of stock market prediction. In recent years, deep learning has achieved tremendous success due to technological advancements such as parallel processing capabilities and big data over the internet. The latest emerging deep learning survey on stock prediction is reported in [5], [6]. Deep learning has been applied to many applications, including time-series problems, particularly stock prediction [7], [8]. When it comes to generic recurrent neural networks (RNNs), the biggest problem is their limited memory, making them unsuitable for retaining longer sequences of data. This difficult challenge is addressed with the help of the long short-term memory (LSTM) network's capabilities [9]. LSTM network architecture is considered a powerful deep learning approach for sequence learning applications such as time series prediction [10].

To the best of our knowledge, most of the study conducted to examine the relationship between news sentiment and stock price movement is in the non-Malaysia context. A deep learning-based technique has been used with sentiment analysis to investigate stock market events and assess the effect on companies from four different nations. The results indicated that using sentiment analysis enhances stock market prediction in Turkey, the US, Pakistan and Hong Kong context [11]. Bidirectional encoder representations from transformer (BERT) were employed to perform sentiment analysis for the stock market movement in a separate investigation in Hong Kong. The findings indicated a 97.35% accuracy in predicting stock price movement [12].

Several studies have proposed deep learning models that incorporate news headlines sentiment analysis to predict stock prices. Zhang *et al*. [13] proposed a novel deep learning model that combines a multi-input convolutional neural network (CNN) and long short-term memory (LSTM) with an attention mechanism to predict stock prices. Their model outperformed traditional machine learning models on real-world stock datasets. Similarly, Wu *et al*. [14] proposed a hybrid model of CNN and LSTM that incorporates normalized news sentiment to predict stock prices. They evaluated their model on the SSECI dataset and demonstrated its effectiveness.

A deep learning model that combines bidirectional LSTM (BiLSTM) with an attention mechanism is proposed by [15] to predict stock prices. Their model incorporates financial news sentiment analysis to capture the sentiment information in the news headlines. The authors tested their proposed model on the S&P 500 dataset and showed its effectiveness. Mudinas *et al*. [16] proposed a model that combines attention-based CNN with LSTM to predict stock prices. Their model utilizes financial news to capture the sentiment information and outperforms traditional machine learning models. Kesavan *et al*. [17] proposed a novel deep learning model that combines LSTM with an attention mechanism and news sentiment analysis to predict stock prices. Their model outperforms traditional machine learning models on the SSECI dataset. Jing *et al*. [18] proposed a deep learning model that combines CNN with LSTM to predict stock prices. Their model incorporates news sentiment analysis to capture the sentiment information in the news headlines. They evaluated their proposed model on the SSECI dataset and showed its effectiveness. Sonkiya *et al*. [19] proposed a multi-granularity news sentiment analysis method for stock price prediction. They evaluated their method on the China A-share market and showed its effectiveness. Similarly, Sachin *et al*. [20] proposed a deep learning model that combines CNN, LSTM, and gated recurrent unit (GRU) with news sentiment analysis to predict the stock prices of the Indian stock market. Their model outperforms traditional machine learning models. Other research that have used deep learning techniques and news sentiment for stock price prediction are reported in [21], [22] for Korean stock market, [23] for China stock market and [24] for New York Stock Exchange.

Although several studies have explored the use of deep learning in combination with news headlines sentiment analysis for stock price prediction, there are limited study on predicting Bursa stock price with Malaysia news headline. One of the major challenges is to improve the accuracy of the models in predicting stock prices for limited news headlines on finance and economic in Malaysia.

Therefore, this research aims to investigate whether news sentiment will affect Bursa Malaysia's stock price movement. The experiment conducted will answer two research questions:
- RQ1: What is the accuracy of the prediction model obtained by combining the Malaysian news sentiment dataset with the Bursa stock price dataset?
- RQ2: Which prediction model, when combined with the sentiment of Malaysian news, is more accurate between LSTM and the decision tree?

The research contribution, therefore, is to explore the accuracy of stock price prediction that incorporates sentiment analysis of Malaysia news headlines, which has not been done before in the context of Bursa stock. By using this approach, we aim to provide an insight in term of the prediction accuracy of stock prices, by comparing two deep learning algorithms namely LSTM and decisition tree. Additionally, the proposed approach can potentially applied to other stock markets and financial instruments.

2. METHOD

The stock prediction model is divided into five steps which are dataset preparation, sentiment analysis, stock prediction model development (LSTM and decision tree (DT)), and evaluation. It begins with the preparation of news sentiment analysis, where the output is later combined with the stock price dataset as a feature for the prediction model as shown in Figure 1. This experiment implements valence aware dictionary for sentiment reasoning (VADER) sentiment intensity analyzer for sentiment analysis with LSTM and DT algorithm for the prediction model. The following subsections discussed each of the steps taken.

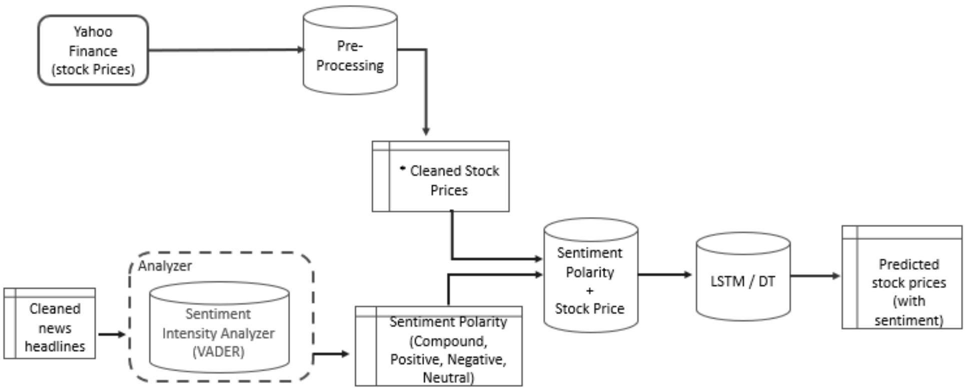

Figure 1. Prediction model proses flow

2.1. Dataset preparation

There are two datasets used in this experiment which are the stock price dataset and financial news headlines. First, the stock price dataset was extracted from five different sectors of blue-chip companies in the Bursa Malaysia stock exchange. A blue-chip stock used in this experiment namely Axiata, Nestle, MISC, Maybank and Sapura Energy is the share of a large company that comes with good financial performance and reputation. Five years of historical stock data were extracted, which comprises seven features which are closing price, high price, low price, adjusted close, date, open price and volume as shown in Figure 2. The five years historical dataset from 2017 to 2022 consists of 7,420 records.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2015-01-02	9.17	9.17	9.05	9.12	5.864488	10217600
1	2015-01-05	9.14	9.17	8.98	9.00	5.787324	14354500
2	2015-01-06	8.90	9.00	8.64	8.80	5.658717	20485400
3	2015-01-07	8.69	8.70	8.58	8.61	5.536541	25785100
4	2015-01-08	8.64	8.76	8.64	8.70	5.594413	13880600

Figure 2. Sample of data extracted from YFinance

Second, is the dataset for news headlines from several finance news portals such as the edge market, business today, financial times, and the new straits times using the Parse hub and leveraging Python BeautifulSoup Package. However, the constraint is the quantity of local news is not sufficient. Therefore, apart from local news, the business news headlines dataset from the Kaggle website is also used as a training dataset for getting an optimal result for sentiment analysis. There are 17,800 news headlines extracted from Kaggle for training purposes and [number] from local news portals as shown in Figure 3.

8-Jan-21	SECTOR AND SUB-SECTOR UPDATE : SUPERCOMNET TECHNOLOGIES BERHAD
23-Dec-20	CHANGE OF NAME: PRESTARIANG BERHAD TO AWANBIRU TECHNOLOGY BERHAD
23-Dec-20	CHANGE OF NAME: REDTONE INTERNATIONAL BERHAD TO REDTONE DIGITAL BERHAD
22-Dec-20	TRANSFER FROM ACE MARKET TO MAIN MARKET: GREATECH TECHNOLOGY BERHAD
	CHANGE OF NAME: GD EXPRESS CARRIER BERHAD TO GDEX BERHAD

Figure 3. Extracted news headlines

2.2. News headlines sentiment analysis

News headlines are analysed using natural language toolkit (NLTK) sentiment VADER for this research. The sentiment analysis tool uses a lexicon and rules-based approach to calculate polarity value to indicate whether the news is positive (>1), negative (<1) or neutral (0) as indicated in Figure 4. Ten news headlines are combined to calculate sentiment for each day.

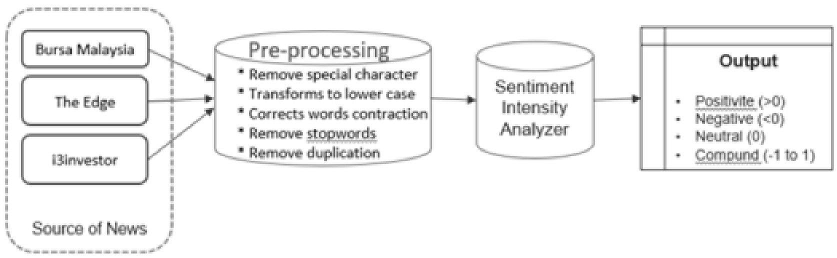

Figure 4. Sentiment analysis proses flow

2.3. Stock price prediction model development

There are two prediction models developed for comparison purposes which are LSTM and DT. The result from sentiment analysis was first combined with the stock price dataset. The sentiment value is appended according to the date of the event. A stock prediction model using LSTM is developed to provide predictions for stock prices. It differentiates recent and previous instances by assigning different weights to the output. As a result, it is more efficient in remembering long sequences of input than other machine learning algorithms, which can only memorize short sequences. The stock price dataset is divided into 70% training data and 30% testing data. The model is developed using the Keras package in Python which is adjusted to 3 epochs and 3 layers of 50 neurons.

The DT model has been used to compare the results to determine whether the accuracy of the LSTM predictions can be justified. Python 3.7.13 was used to write and evaluate the models. The scikit-learn package was used to import the decision tree classifier and assessment metrics. This is considered a sub-experiment that employs two set of datasets from the New York stock exchange (NYSE) and Bursa Malaysia that were crawled using the same technique as mentioned in the previous chapter.

2.4. Evaluation

The root mean square error (RMSE) is used to evaluate the efficacy of the model. It is calculated by comparing the difference between each observed (close price) and predicted value. The use of RMSE is quite common and considered an excellent general error metric for numerical predictions. The smaller the RMSE value, the closer the predicted and observed values are [25]. Therefore, if the model's objective is to predict daily stock prices (time series problems), it will be assessed using the RMSE approach. To avoid overfitting, which neural networks are generally prone to, dropout layers are continually inserted after each LSTM layer.

3. RESULTS AND DISCUSSION

3.1. Prediction using LSTM model (RQ1)

Stock market or stock prices are an example of a time series problem and investing in the stock market is an important part of the financial sector. The prediction of stock price fluctuation behaviour is relying on many factors and parameters such as historical stock price data. Therefore, to produce predictions that are as accurate as they possibly can be, this research project is making use of these historical stock prices as input to the LSTM prediction model.

In this experiment, the LSTM prediction model is developed using Keras as a deep learning framework. Additionally, five years of historical stock prices are used as the input for the model. After ensuring that the prediction model is consistent with the training data, only then testing data is used for real prediction. The result in Table 1 is the answer to RQ1.

Table 1. RMSE for LSTM on Bursa Malaysia stock exchange

Company	Business sector	RMSE (LSTM)
Axiata Group Berhad	Telecommunications	0.050269
Nestle Berhad	Consumer Products	0.769993
Maybank Berhad	Banking	0.056826
MISC Berhad	Logistics	0.075765
Sapura Energy Berhad	Energy	0.042952

Table 1 is showing the RMSE for all the selected companies. These RMSE values are different from one company to another due to the consistency and variation of the supplied dataset. In this experiment, some training analysis using LSTM model, after running numerous training analyses using the LSTM model, it is found that 3 epochs with 3 hidden layers and 1 dense layer produced the best results. The RMSE is recorded significantly lower and it is indicating that the accuracy is higher. The experiment results indicate that deep learning algorithms have a tremendous impact on the financial sector, particularly in terms of developing time-series-based prediction models. They outperform all other regression models in terms of accuracy when used to predict stock price and it conforms to what has proposed previously [26].

As indicated in Table 1, five different business sectors recorded significantly low RMSE values. Sapura energy berhad records the lowest RMSE value (0.042952) among the others after 3 epochs of model fitting. While Nestle Berhad records the highest RMSE value (0.769993). During the process of constructing the prediction model, the optimizer Adam is utilised so that the data dropout may be managed effectively. The model begins to drastically overfit the data when the number of epochs used in the training is increased. As a result, the number of epochs will need to be modified following the validation loss. As stated in the research objective, the primary concern was to evaluate how well the LSTM model could predict the stock market. For this purpose, there is a broad variety of statistical methods that can be used to quantify this accuracy, but this research project relied on the RMSE value that was produced by the prediction model as it is frequently used by the data science community to measure the accuracy errors.

3.2. Comparison between LSTM and DT (RQ2)

The DT model has been used to compare the results to determine whether the accuracy of the LSTM predictions can be justified. Python 3.7.13 was used to write and evaluate the models. The scikit-learn package was used to import the decision tree classifier and assessment metrics. This is considered a sub-experiment that employs two sets of datasets from the NYSE and Bursa Malaysia that were crawled using the same technique as mentioned in the previous chapter. Table 2 shows the result for answering RQ2.

Table 2. RMSE for LSTM and DT on Bursa Malaysia and NYSE

Stock exchange	Company	Business sector	RMSE (LSTM)	RMSE (DT)
Bursa Malaysia	Sapura Energy Berhad	Energy	0.042952	0.063845
	Axiata Group Berhad	Telecommunications	0.050269	0.051586
	Nestle Berhad	Consumer Products	0.769993	0.915638
	Maybank Berhad	Banking	0.056826	0.064698
	MISC Berhad	Logistics	0.075765	0.087364
NYSE	Amazon Inc	Internet Retail	10.550978	29.568602
	Alphabet Inc	Communication Services	4.046784	16.677339
	Microsoft Corporation	Software Infrastructure	0.225494	2.040376
	Fortinet Inc	Technology	0.378289	2.996243
	Palo Alto Network, Inc	Technology	0.113574	4.148382

Knowing that the information technology industry dominates 27.6% of the NYSE stock market in terms of market capitalization, five different technology companies' datasets have been crawled as input to the prediction models (LSTM and DT). The entire market value sets a record of $95 trillion in November 2020, surpassing the levels seen before the coronavirus outbreak. These inject volatility and variance into the datasets, which may result in a different RMSE value being produced by these two algorithms. The results are compared and presented in Table 2 further in this experiment.

3.3. Discussion on models comparison

The decision tree algorithm is far less challenging and not complicated to build, it does not require a huge dataset and can be trained in less time than other models. That makes a decision tree a popular algorithm to solve supervised machine learning problems. The most challenging area of machine learning is developing deep learning algorithms such as the LSTM algorithm. Normally, deep learning models require huge training data and time, as a result, they often produce greater prediction accuracy as deep learning models perform automated feature extraction and classification running simultaneously [27]. This contrasts with machine learning algorithms, which require a feature selection process before training. Deep learning models can perform automated feature extraction and classification concurrently.

Referring to Table 2 indicates the comparison of RMSE values for LSTM and DT for 2 different datasets. The first part of the table (in green) represents the dataset from Bursa Malaysia while the dataset from the NYSE is represented in the second part of the table (in yellow). For all five different business sectors in Bursa Malaysia, LSTM outperforms DT by a slim difference. For instance, the LSTM and DT both produce the highest RMSE values for Nestle Berhad, which give 0.769993 and 0.915638 respectively. This makes the highest difference between these two algorithms with a 0.145645 value in RMSE. While the lowest difference in RMSE value between LSTM and DT is 0.0012896 for Axiata Group Berhad whereby LSTM and DT generate 0.050269 and 0.051586 respectively. The difference in RMSE between these two algorithms is not significant since neither the highest nor the lowest results are not even reaching 1.

In the second part of Table 2, LSTM continues to perform better than DT for each of the five companies that were chosen, with a significant difference in RMSE. LSTM produces a small RMSE value rather than DT for Amazon Inc company with 10.550978 and 29.568602 respectively. This leads to the highest difference in RMSE for both algorithms in this experiment, which uses the NYSE dataset and is considered significant enough in comparing between models.

The lowest difference in RMSE is 1.814882 for Microsoft Corporation whereby LSTM and DT produce 0.225494 and 2.040376 respectively. The LSTM performs significantly better than DT for each of the selected companies in the NYSE dataset. When a company's stock prices are high, this might lead to the relative RMSE being high as well.

3.4. Discussion on dataset comparison

Referring to Table 2, it is evident that the DT algorithm produces higher RMSE values than LSTM, particularly for the NYSE dataset; Amazon Inc stock price, where the DT algorithm produces 29.568602 whereas LSTM gives 10.550978. When it comes to making predictions regarding stock prices, this is regarded as a significant difference. This occurred as a direct result of the significant gap in magnitude between closing prices for a certain period. It is proves that Figure 5 shows the fluctuation of the closing price and the magnitude is very high. For instance, from 12th July 2021 to 29th April 2022 (less than a year) the magnitude of price changes is more than USD 1,100 per unit share.

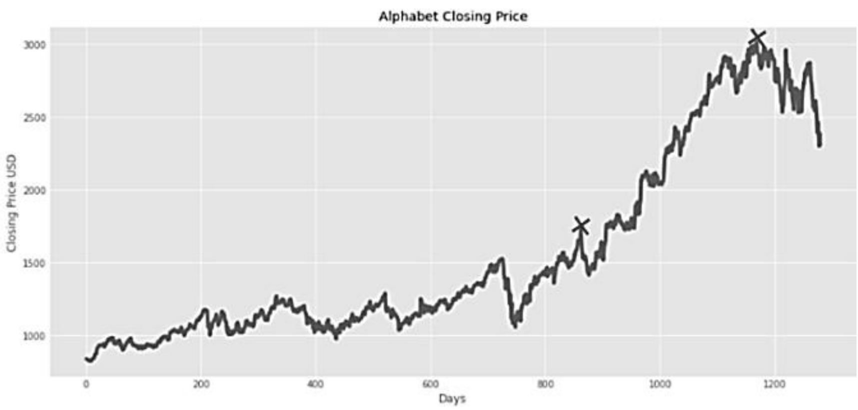

Figure 5. Alphabet Inc closing price

Another case from NYSE dataset that may be used as concrete evidence is that Alphabet stock prices, as shown in Table 2, indicate that DT and LSTM give values of 16.677339 and 4.046784 in RMSE respectively. For instance, between the 18th of November 2020 and the same date the following year, the magnitude of the closing price is more than USD 1,200 per unit share as in Figure 6 When comparing the stock prices of Amazon and Alphabet, it is found that both companies have a greater magnitude of closing prices, and also the stock prices of both companies are higher. These elements (the stock price and the magnitude of the closing price) need to be in a higher value for them to influence the RMSE value.

Figure 6. Amazon closing price

4. CONCLUSION

Based on the results of this experiment, which utilized two distinct datasets, it can be concluded that the LSTM deep learning algorithm is superior to DT for predicting stock prices. This finding reinforces the notion that choosing a deep learning algorithm is the most effective approach for stock price prediction. With the advancements in deep learning and increased computing power, various types of machine and deep learning models have been developed and tested for stock prediction. This project has made contributions to three distinct areas related to investment and trading in Malaysia. The first area involves the sentiment analysis of local news headlines. This analysis is important because it can help determine the public's overall sentiment and can be used as a leading indicator for investment decisions. The second area of focus was the development of a stock price prediction model. This model was designed to assist investors and traders in predicting future stock prices, enabling them to earn higher profits. Finally, the project team integrated Malaysia news sentiments into Bursa stock price prediction. This integration of news sentiments and closing stock prices can aid investors in making more informed investment decisions by providing additional indicators to consider before making recommendations. Overall, these three areas of focus have the potential to significantly benefit investors in Malaysia.

ACKNOWLEDGEMENT

The authors express their sincere appreciation to the reviewers for their valuable feedback, which helped enhance the quality of this article. The funding for this research was provided by 'Skim Zamalah UTeM' of Universiti Teknikal Malaysia Melaka (UTeM), Malaysia.

REFERENCES

[1] N. Sharma and A. Juneja, "Combining of random forest estimates using LSboost for stock market index prediction," in *2017 2nd International Conference for Convergence in Technology, I2CT 2017*, Apr. 2017, vol. 2017-Janua, pp. 1199–1202, doi: 10.1109/I2CT.2017.8226316.

[2] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "Predicting stock market price movement using sentiment analysis: Evidence from ghana," *Applied Computer Systems*, vol. 25, no. 1, pp. 33–42, May 2020, doi: 10.2478/acss-2020-0004.

[3] W. Long, L. Song, and Y. Tian, "A new graphic kernel method of stock price trend prediction based on financial news semantic and structural similarity," *Expert Systems with Applications*, vol. 118, pp. 411–424, Mar. 2019, doi: 10.1016/j.eswa.2018.10.008.

[4] M. R. Islam, I. F. Al-Shaikhli, R. B. M. Nor, and V. Varadarajan, "Technical approach in text mining for stock market prediction: A systematic review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 2, pp. 770–777, May 2018, doi: 10.11591/ijeecs.v10.i2.pp770-777.

[5] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," *Expert Systems with Applications*, vol. 184, p. 115537, Dec. 2021, doi: 10.1016/j.eswa.2021.115537.
[6] Z. Hu, Y. Zhao, and M. Khushi, "A survey of forex and stock price prediction using deep learning," *Applied System Innovation*, vol. 4, no. 1, pp. 1–30, Feb. 2021, doi: 10.3390/ASI4010009.
[7] W. Jiang and L. Zhang, "Edge-SiamNet and Edge-TripleNet: New deep learning models for handwritten numeral recognition," *IEICE Transactions on Information and Systems*, vol. E103D, no. 3, pp. 720–723, Mar. 2020, doi: 10.1587/transinf.2019EDL8199.
[8] H. K. Sagiraju and S. Mogalla, "Application of multilayer perceptron to deep reinforcement learning for stock market trading and analysis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 3, pp. 1759–1771, Dec. 2021, doi: 10.11591/ijeecs.v24.i3.pp1759-1771.
[9] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.
[10] M. Pacella and G. Papadia, "Evaluation of deep learning with long short-term memory networks for time series forecasting in supply chain management," *Procedia CIRP*, vol. 99, pp. 604–609, 2021, doi: 10.1016/j.procir.2021.03.081.
[11] H. Maqsood *et al.*, "A local and global event sentiment based efficient stock exchange forecasting using deep learning," *International Journal of Information Management*, vol. 50, pp. 432–451, Feb. 2020, doi: 10.1016/j.ijinfomgt.2019.07.011.
[12] M. Li, W. Li, F. Wang, X. Jia, and G. Rui, "Applying BERT to analyze investor sentiment in stock market," *Neural Computing and Applications*, vol. 33, no. 10, pp. 4663–4676, May 2021, doi: 10.1007/s00521-020-05411-7.
[13] H. Zhang *et al.* "Leveraging statistical information in fine-grained financial sentiment analysis," *World Wide Web*, vol. 25, no. 2, pp. 513–531, 2022, doi: 10.1007/s11280-021-00993-1.
[14] S. Wu, Y. Liu, Z. Zou, and T. H. Weng, "S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis," *Connection Science*, vol. 34, no. 1, pp. 44-62, Jun. 2021, doi: 10.1080/09540091.2021.1940101.
[15] S. L. O. Lim, H. M. Lim, E. K. Tan, and T. P. Tan. "Examining machine learning techniques in business news headline sentiment analysis," in *Computational Science and Technology: 6th ICCST 2019*, 2019. pp. 363-372, doi: 10.1007/978-981-15-1676-2_34.
[16] A. Mudinas, D. Zhang, and M. Levene. "Market trend prediction using sentiment analysis: lessons learned and paths forward,". 2019, *arXiv:1903.05440*.
[17] M. Kesavan, J. Karthiraman, R. T. Ebenezer, and S. Adhithyan. "Stock market prediction with historical time series data and sentimental analysis of social media data," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 477-482, doi: 10.1109/ICICCS48267.2020.9120255.
[18] N. Jing, Z. Wu, and H. Wang. "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Systems with Applications*, vol. 178, 2021, doi: 10.1016/j.eswa.2021.115019.
[19] P. Sonkiya, V. Bajpai, and A. Bansal. "Stock price prediction using BERT and GAN," 2021, *arXiv:2107.09055*.
[20] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, and P. Nagrath. "Sentiment Analysis Using Gated Recurrent Neural Networks," *SN Computer Science*, vol. 74, 2020, doi: 10.1007/s42979-020-0076-y.
[21] Y. Xu and V. Keselj, "Stock Prediction using Deep Learning and Sentiment Analysis," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 5573-5580, doi: 10.1109/BigData47090.2019.9006342.
[22] K. Nam and N. Seong, "Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market," *Decision Support Systems*, vol 117, pp. 100-112, Feb. 2019, doi: 10.1016/j.dss.2018.11.004.
[23] J. Long, Z. Chen, W. He, T. Wu, and J. RenJ, "An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market," *Applied Soft Computing*, vol 91, Jun. 2020, doi: 10.1016/j.asoc.2020.106205.
[24] 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*, vol 13, pp. 3433–3456, 2020, doi: 10.1007/s12652-020-01839-w.
[25] A. H. Suyanto, T. Djatna, and S. H. Wijaya, "Mapping and predicting research trends in international journal publications using graph and topic modeling," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 30, no. 2, pp. 1201–1213, May 2023, doi: 10.11591/ijeecs.v30.i2.pp1201-1213.
[26] M. A. I. Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using LSTM and Bi-directional LSTM Model," in *2nd Novel Intelligent and Leading Emerging Sciences Conference, NILES 2020*, Oct. 2020, pp. 87–92, doi: 10.1109/NILES50944.2020.9257950.
[27] C. N. Kamath, S. S. Bukhari, and A. Dengel, "Comparative study between traditional machine learning and deep learning approaches for text classification," in *Proceedings of the ACM Symposium on Document Engineering 2018, DocEng 2018*, Aug. 2018, pp. 1–11, doi: 10.1145/3209280.3209526.

BIOGRAPHIES OF AUTHORS

Zuleaizal bin Sidek is the founder and managing director of Kencana Niaga, an IT start-up and independent provider of data analytics for interdisciplinary analysis. He is also a postgraduate student in Philosophy Doctor in Data Science at the Universiti Teknikal Malaysia (UTeM). He is currently working as Data Scientist at the Institut Tun Perak (Melaka state owned company). He has multiple experiences in IT for more than 10 years at the Universiti Teknologi MARA as an IT Manager. His first paper on e-government services: A value of E-Service in Local Government: A fuzzy approach evaluation". His areas of interest are big data management and analytics, machine learning and blockchain technology. He can be contacted at email: zulsidek@gmail.com.

Assoc. Prof. Dr Sharifah Sakinah Syed Ahmad is currently an associate professor in the Department of Intelligent Computing and Analytics (ICA), Faculty of Information & Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM). She received her bachelor's and master's degrees in applied mathematics from the School of Mathematics at the University of Science, Malaysia. Following this, she received her Ph.D. from the University of Alberta, Canada in 2012 in intelligent systems. She can be contacted at email: sakinah@utem.edu.my.

Dr Noor Hasimah Ibrahim Teo is a senior lecturer at the Universiti Teknologi Teknologi MARA Melaka, Malaysia. She received her Ph.D. in Computer Science from the University of Warwick, United Kingdom in the year 2019 and an M.Sc. in Computer Science from Universiti Teknologi MARA, Malaysia. She received Ernst Mach Grant ASEA-UNINET, OeAD Austria for a postdoctoral visit to the Technological University of Vienna in 2022. Her research interests are concentrated in the field of ontology, text processing, machine learning, and question answering system. She can be contacted at email: shimateo@uitm.edu.my.