

Time Series Analysis on Forex Exchange using Artificial Neural Network and K-Nearest Neighbor Algorithms

Fong Shu Ying¹, Norhaidah Mohd Asrah^{1*}

¹ Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 86400 Pagoh, Muar, Johor, MALAYSIA.

*Corresponding Author: norhaida@uthm.edu.my

DOI: <https://doi.org/10.30880/ekst.2024.04.02.037>

Article Info

Received: 27 December 2023

Accepted: 11 January 2024

Available online: 12 December 2024

Keywords

Artificial Neural Network, K-Nearest Neighbor, Forex Exchange

Abstract

Accurate forex rate prediction is crucial for investors and businesses looking to optimize returns and manage their investment portfolios sensibly in the ever-changing financial markets. With its frequent fluctuations, interdependence, and the influence of world economic events, the forex market presents certain inherent difficulties that are the subject of this study. Given these difficulties, the study focuses on developing and application of reliable forecasting models and approaches. The methodology involves the construction and comparison of Artificial Neural Network (ANN) and K-Nearest Neighbor (KNN) models, with the objective of selecting the most accurate model for subsequent multi-step ahead predictions. Results indicate that KNN models outperform ANN models and are chosen for making 30-day ahead predictions. For short term forecasting, KNN's forecasts are valid, but there are issues when looking farther than a week ahead. The study's conclusion highlights how challenging long-term forecasting in the forex market is by its own nature of chaotic and noisy. The results highlight how crucial it is to keep up research and innovate in order to improve forecasting models in this complex financial environment.

1. Introduction

Foreign exchange, or forex, is the term used to describe the activity of exchanging currencies, which is the buying and selling of various currencies to profit from changes in their relative prices [1]. It is a crucial component of international trade, investment, and tourism. Currency fluctuations can be attributed to a variety of factors, including market expectations, macroeconomic conditions, and technological advancements that enable traders to obtain information quickly. Majors, or the most frequently traded currencies, cross-rates, or exchange rates not based on the US dollar, and exotics, or minor currencies, make up the forex market. Because it is the world's reserve currency and is impacted by political decisions, economic data, and the US's influence on the world economy, the USD is heavily weighted in the forex market [2]. The Singapore Dollar (SGD) and Malaysian Ringgit (MYR) are also important currencies in the foreign exchange market. The SGD is known for its strength and stability, while the MYR is impacted by Malaysia's political stability, economic performance, and central bank policies [3], [4].

Forecasting exchange rates for the USD, SGD, and MYR are essential for managing investment portfolios and maximizing profits. However, because traditional time series models are not suitable for the non-linear chaotic and volatile nature of the forex market, more advanced forecasting models like machine learning approaches are more suitable [5], [6], [7], [8]. Among all machine learning algorithms, the K-nearest neighbor (KNN) algorithm stands out as one of the most widely used machine learning models in forex forecasting. KNN algorithm is a non-

This is an open access article under the CC BY-NC-SA 4.0 license.



parametric supervised learning model, learns iteratively from a training dataset, allowing it to make predictions based on acquired knowledge [9]. Researches indicate that KNN models exhibit superior prediction accuracy in forex compared to traditional time series models [10], [11]. However, a limitation of the KNN algorithm lies in choosing the appropriate k-value. A large k-value may introduce more bias and overly smoothed forecasts, while a small k-value could lead to overfitting and increased susceptibility to noise [12].

Other than KNN algorithm, the researchers at the time preferred Artificial Neural Network (ANN) algorithm as its neural-network based approach is capable in learning from big data, autonomously extracting features, adapting to changing settings, and address issues across numerous industries with high accuracy and robustness [13], [14], [15]. ANN is a machine learning algorithm inspired by the anatomy and functioning of the human brain, comprising multiple layers of interconnected nodes or neurons that process and transmit data through a network of weighted connections. ANN are applied in various of field, especially in financial market, as many researches showed that ANN is possible to predict chances in stock price with high volatility as it can process the data and identify patterns that may be difficult for humans to detect [16], [17], [18].

This study aims to construct forecasting models based on the exchange rate of USD/MYR and SGD/MYR using ANN and KNN algorithms, compare the performance of ANN and KNN models using mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE) and coefficient of determination (R^2) and finally forecast the exchange rate using the best model.

2. Methodology

2.1 Data Description

The forex dataset in this study was access from the Investing.com. There are 2 datasets used and both of them were retrieved from <https://www.investing.com/currencies/sgd-myr-historical-data> and <https://www.investing.com/currencies/usd-myr-historical-data> via Investing.com. The USD/MYR dataset consists of 4552 daily price data while the SGD/MYR dataset consists of 4584 daily price data. The period of the datasets was started from January 2006 to August 2023.

From Fig. 1, the USD/MYR prices declined from the beginning of 2006 until September 2008. Subsequently, they experienced an upward trend until December 2008, followed by another decline until September 2011. The lowest recorded price occurred on August 1, 2011. Prices continued to fluctuate between September 2011 and October 2014, followed by a sharp upward trend until October 2015. Afterward, prices fluctuated between 3.8 and 4.5, reaching a peak at 4.746 on November 4, 2022, before decreasing.

Based on Fig. 2, it is evident that SGD/MYR prices exhibited a steady increase from January 2006 to November 2006, followed by a decreasing trend until September 2007, with the lowest recorded price of 2.210 on May 22, 2007. Then, the prices began to rise steadily, followed by a sharp fall in March 2010. Afterward, the prices increased gradually with some fluctuations until June 2015, when they suddenly spiked and surpassed the 3.0 exchange rate in August 2015. After that, the prices fluctuated with both rises and falls until September 2022. In October 2022, prices started to rise again, reaching a peak on July 12, 2023.

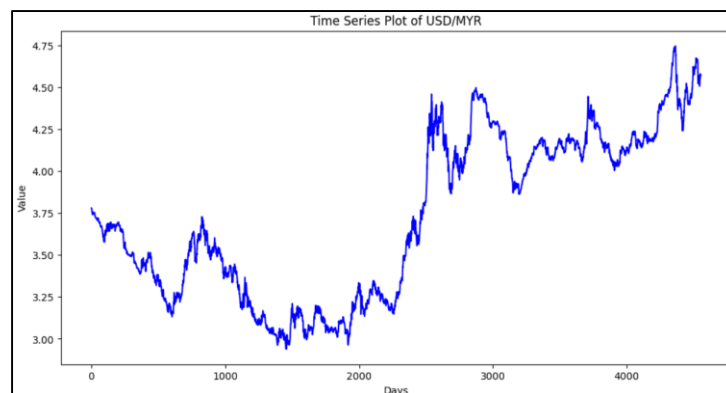


Fig. 1 Time Series Plot of USD/MYR from 1st January 2006 until 31st July 2023

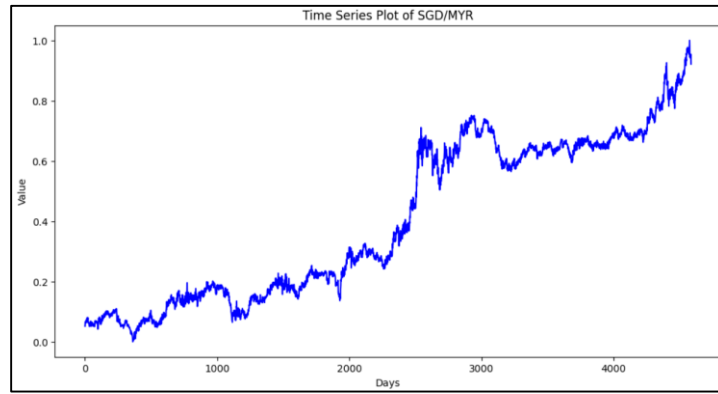


Fig. 2 Time Series Plot of SGD/MYR from 1st January 2006 until 31st July 2023

2.2 Data Transformation

Min-max transformation was applied to reduce the range of the data. In this technique, the data values were subtracted with the minimum value and divided by the range. The transformation process can be written as equation (1), where X_i is the new x value, x_i is the actual value, x_{min} is the minimum value in the data, x_{max} is the maximum value in the data and $I = 1, 2, 3, \dots, n$ [19].

$$X_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

2.3 K-Nearest Neighbor (KNN)

KNN is a frequently used method in forecasting study, especially in predicting financial products. The regression model of KNN is differ from the classification model of KNN, as it maps patterns to continuous labels, to make prediction through computed the mean of function values [12]. The formula of the functional regression model of KNN is as equation (2), where Q_i is the forecast value, w_k is the weight of the observed value, x_k is the observed value and k is the number of nearest neighbors [20].

$$Q_i = \sum_{k=1}^K (w_k \times x_k) \tag{2}$$

2.4 Artificial Neural Network (ANN)

In 1943, McCulloch & Pitts introduced ANN and it duplicates the process of conveying and processing data through learning, simulating how a human brain neuron function. Fig. 3 shows how neural networks mimic the organization of human brain neurons, where dendrites receive information and transmit it to the nucleus through axons [21]. Axon terminals will get the information after the nucleus has processed it.

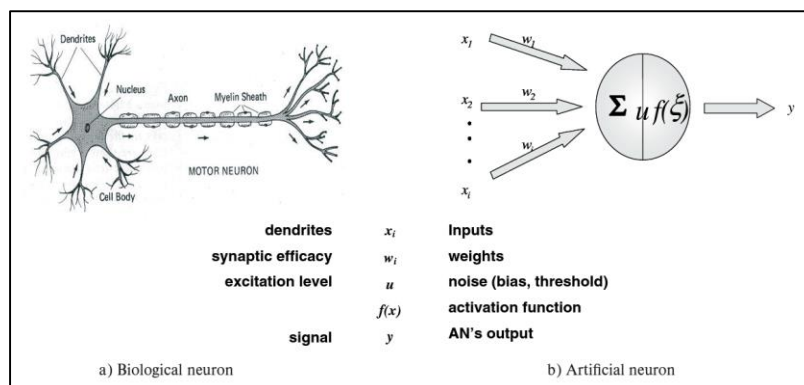


Fig. 3 Comparison between biological neuron and artificial neuron

The ANN model used in this study is feedforward Multi-Layer Perceptron (MLP) network and trained via back propagation. The reason to choose back propagation as training algorithms is research shows that if given sufficient free parameters, a typical back propagation ANN got the ability of approximate any function and Fig. 4 illustrates how the MLP model works.

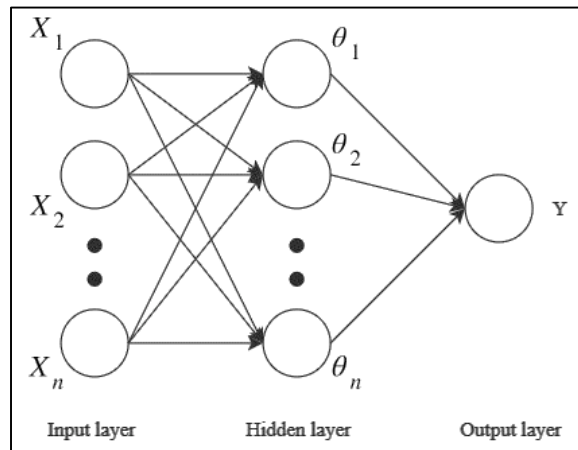


Fig. 4 Architecture of MLP model

The input to the hidden neurons in hidden layer can be written as equation (3) and input to the output neurons can be written as equation (4), where $(H_{input})_n$ is the i th input to the n th hidden neuron, I_i is the value at the i th input neuron, W_{in} is the connection weight from the i th input neuron to the n th hidden neuron, θ_n is the threshold of n th hidden neuron, p is the number of input neurons, q is the number of hidden neurons, $(O_{input})_m$ is the input of $(H_{output})_n$ to the m th hidden neuron, $(H_{output})_n$ is the n th hidden neuron output, W_{nm} is the interconnection weight between the n th hidden neuron and m th output neurons, θ_m is the threshold of m th output neuron, r is the number of output neurons, $n = 1, 2, 3, \dots, q$ and $m = 1, 2, 3, \dots, r$ [23].

$$(H_{input})_n = \sum_{i=1}^p I_i \times W_{in} - \theta_n \tag{3}$$

$$(O_{input})_m = \sum_{n=1}^q (H_{output})_n \times W_{nm} - \theta_m \tag{4}$$

The equation of the outputs in hidden and output layer is as equation (5) and equation (6) respectively, where H_j is the output of j th hidden layer, $f()$ is the transfer function, x_i is the i th input vector, y_m is the output value of m th nodes in the output layer and $j = 1, 2, 3, \dots, m$ [24].

$$y_m = f \left[\sum_{j=1}^m W_{nm} H_j + \theta_m \right] \tag{5}$$

$$H_j = f \left[\sum_{i=1}^p W_{in} x_i + \theta_n \right] \tag{6}$$

The activation functions, $f(x)$ is made to accommodate the non-linearity relationship between input and output. One of the widely used activation function is the sigmoid transfer function in equation (7).

$$f(x) = sig(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

2.5 Model Parameter Settings

The training parameter of the ANN models are set to have 1 input layer, 1 hidden layer and 1 output layer. The activation functions set for input and hidden layer is sigmoid transfer function while the output layer is activated by linear transfer function. There were 64 neurons each in the input and hidden layer but only 1 neuron in the output layer. The models are trained in a batch size of 16, 10000 epochs and learning of 0.01 unit. All the ANN models were set in the same settings for both USD/MYR and SGD/MYR forecast. In the other side, the number of nearest neighbors of KNN models are set as 225 while training with both datasets. 70% of the datasets were used to construct the forecasting models, while the remaining 30% of the data were served as the testing data for the model evaluation.

2.6 Forecast Accuracy Measurement

The parameters used to evaluate the performance of the models are mean absolute error (MAE) in equation (8), mean squared error (MSE) in equation (9), root mean square error (RMSE) in equation (10) and coefficient of determination (R^2) in equation (11), where y_t is the actual value at time t , y_{t-1} is the observed value at time $t - 1$, \hat{y}_t is the predicted value at time t and n is the number of observations [25]. MAE measures how close the actual value and predicted value are, while MSE measures mean squared differences of the actual value and predicted value. RMSE is the standard deviation of the predicted errors and R^2 measures how well the independent variable explains the variation in the dependent variable. Lower values from MAE, MSE and RMSE indicate fewer errors produced while higher value from R^2 suggests that the predictions data fit well in the actual data. Thus, the forecasting model with lowest values in the error measurements and highest value of R^2 will be considered as the best accuracy model.

$$MAE = \frac{\sum_{i=1}^n |y_t - \hat{y}_t|}{n} \quad (8)$$

$$MSE = \frac{1}{n^2} \sum_{i=1}^n (y_t - \hat{y}_t)^2 \quad (9)$$

$$RMSE = \frac{1}{n^2} \sqrt{\sum_{i=1}^n (y_t - \hat{y}_t)^2} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_t - \hat{y}_t)^2}{\sum_{i=1}^n (y_t - \bar{y}_t)^2} \quad (11)$$

2.7 Multi-steps Ahead Forecast

Various type of multi-steps ahead forecasting methods is found in forecasting. In this study, the Direct strategy was used to make multi-steps ahead forecast as this strategy is immune to the accumulation of errors because no approximated values were used to compute the forecasts [25]. The equation for the multi-steps ahead forecasting is as equation (12), where y_{t+l} is the multi-steps ahead forecast at time $t + l$, x_t is the forecast value at time t , l is the number of timesteps, and w is the errors term, included noise, modeling errors.

$$y_t = f(x_1, x_2, x_3, \dots, x_t) + w \quad (12)$$

3. Results and Discussion

3.1 USD/MYR Dataset

In Fig. 5, the actual versus predicted values are illustrated for the ANN and KNN models on the testing data of USD/MYR dataset. it is evident that the ANN model's predictions align closely with the actual values before

reaching 1000th day. However, beyond that point, the forecasted values begin to deviate significantly from the actual data. In comparison, the KNN model appears to exhibit superior predictive accuracy. As seen in Fig. 5, the predictions generated by the KNN model closely match the actual values.

When considering the measures of errors in Table 1, it is evident that the KNN model outperforms the ANN model, boasting the lowest values for MAE, MSE, and RMSE. This implies that the KNN model exhibited superior accuracy, with fewer errors in its predictions compared to the ANN model. Furthermore, the KNN model showcases a higher R^2 value when contrasted with the ANN model. This indicates a stronger alignment between the KNN model's predictions and the actual values. As a result, the KNN model achieved a higher level of prediction accuracy in the USD/MYR dataset when compared to the ANN model.

Table 1 : Measures of errors for USD/MYR dataset forecasting

	MAE	MSE	RMSE	R^2
ANN	0.05191327	0.00557077	0.07463759	0.92444207
KNN	0.01591597	0.00043851	0.02094074	0.99381205

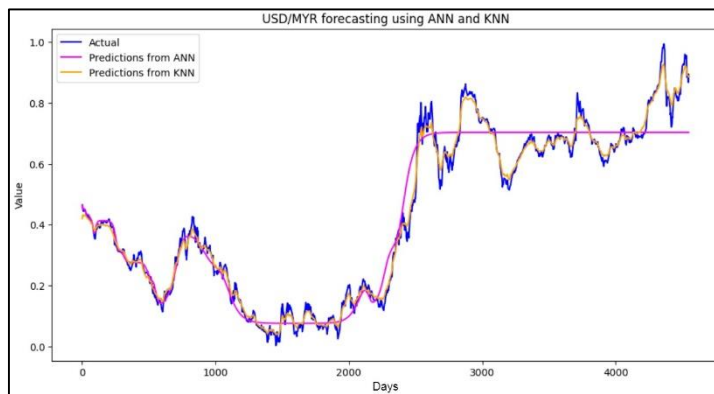


Fig. 5 Actual versus prediction of USD/MYR forecasting using ANN and KNN

3.2 SGD/MYR Dataset

When examining the SGD/MYR prediction in Fig. 6, it becomes apparent that the ANN model effectively captures the overall trend of the data. However, it struggles to predict the data's fluctuations accurately. In contrast, the KNN model once again demonstrates strong predictive capabilities, as evidenced in Fig. 6.

The assessment of error metrics presented in Table 2 further corroborates the KNN model's superior performance. The model exhibits the lowest values for MAE, MSE, and RMSE, as well as the highest R^2 value. Consequently, the KNN model emerges as the top-performing model among the two datasets when compared to the ANN model.

Table 2 : Measures of errors for SGD/MYR forecasting

	MAE	MSE	RMSE	R^2
ANN	0.00391619	0.04306639	0.06257943	0.94623111
KNN	0.01046194	0.00021329	0.01046194	0.99705487



Fig. 6 Actual versus prediction of SGD/MYR forecasting using ANN and KNN

3.3 Multi-steps Ahead Forecast

Since KNN model has better performance than ANN, KNN forecast models were employed to generate forecasts for USD/MYR and SGD/MYR, predicting 30 days ahead. Fig. 7 and Fig. 8 depict the forecasts from KNN models in comparison to the actual values for USD/MYR and SGD/MYR predictions. Drawing insights from Fig. 7, the predictions generated by the KNN model exhibit remarkable accuracy up to the fourth day. However, beyond this point, the KNN model's predictions appear to deviate from the actual values. Despite this, the model still manages to capture the trend of USD/MYR prices until the twelfth day. Subsequently, the predictions start to significantly diverge from the actual prices. The KNN model forecasts a decreasing trend after the twelfth day, whereas the actual prices exhibit an upward trajectory with some fluctuations.

Based on Fig. 8, up until the fifth day, the KNN model perfectly predicts SGD/MYR prices. After this, the model predicts a declining trend, which is at odds with the actual prices, which show fluctuations and an upward trend.

By analyzing the outcomes, the predictions crafted by the KNN models prove promising for short-term forecasting but fall short in validity for long-term forecasting. The accuracy of KNN model predictions remains robust for approximately 4 to 5 days. Beyond this initial period, the model retains its ability to forecast the trend of forex data for around twelve days, after which the predictions become increasingly unreliable.

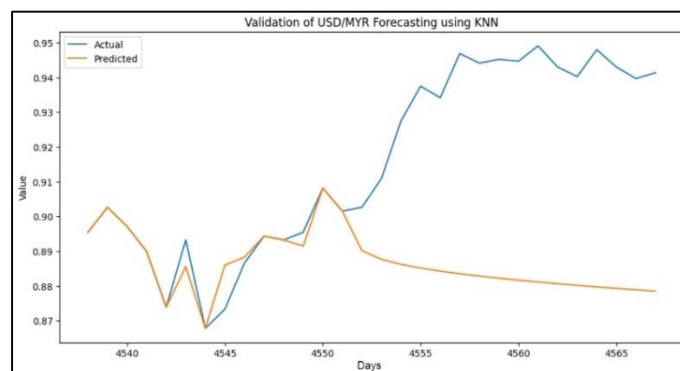


Fig. 7 Validation of USD/MYR Forecasting using KNN

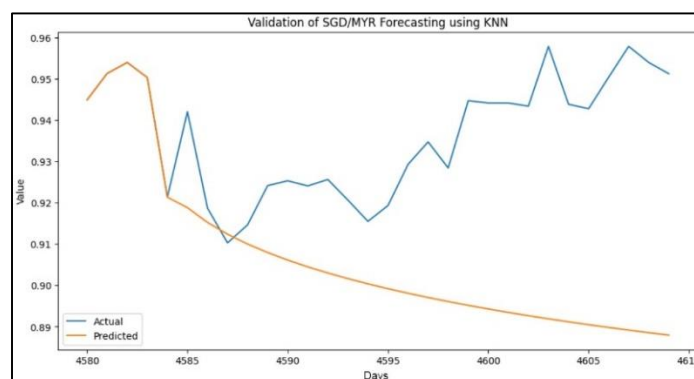


Fig. 8 Validation of SGD/MYR Forecasting Using KNN

4. Conclusion

This study successfully constructed forecasting models based on ANN and KNN algorithms by utilizing 70% of USD/MYR and SGD/MYR data as training data to train the ANN and KNN models. The training parameters for the ANN models were set with 1 input and 1 hidden layer, each containing 64 neurons, and 1 output layer with 1 neuron. The ANN models were trained with a batch size of 16 over 10,000 epochs with a learning rate of 0.01 unit. Meanwhile, the KNN models were configured with 225 nearest neighbors. All ANN and KNN models were set with identical configurations to ensure fair and consistent results.

Besides, this study managed to choose the best accuracy model by the comparison of compare the performances of ANN and KNN models. Performance measurements in this study included mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and R^2 . The performance of KNN models in USD/MYR and SGD/MYR datasets was proven to be superior compared to the ANN models. KNN models exhibited the lowest values for MAE, MSE, and RMSE, and the highest values for R^2 . Based on the time series plots in Fig. 5 and Fig. 6, the performance of KNN models closely aligns with the actual testing data. Therefore, the KNN model is concluded as the superior model compared to the ANN model.

Lastly, the KNN models were employed to make a 30-day ahead forecast for USD/MYR and SGD/MYR. The predictions from KNN models proved to be overwhelming for short term prediction, but inaccurate for long term prediction as the results did not align with the actual data. Additionally, the results demonstrate that long-term forecasting is challenging for forex data due to its chaotic nature and the presence of noise.

In conclusion, using different machine learning-based methods, like Radial Basis Function Networks (RBFNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory Networks (LSTMs), can improve the accuracy of forex price forecasting. It is advised to conduct more research on the best network architectures for increased accuracy. To account for all influencing factors, a hybrid model that integrates quantitative and qualitative forecasting methods is advised. This comprehensive model reduces the possibility of unfavorable events, which eventually results in more precise forex price forecasts.

Acknowledgement

The authors would thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: **study conception and design:** Fong Shu Ying, Norhaidah Mohd Asrah; **data collection:** Fong Shu Ying; **analysis and interpretation of results:** Fong Shu Ying, Norhaidah Mohd Asrah; **draft manuscript preparation:** Fong Shu Ying, Norhaidah Mohd Asrah. All authors reviewed the results and approved the final version of the manuscript.

References

- [1] Sarangi, P. K., Chawla, M., Ghosh, P., Singh, S., & Singh, P. K. (2022). FOREX trend analysis using machine learning techniques: INR vs USD currency exchange rate using ANN-GA hybrid approach. *Materials Today: Proceedings*, 49, pp. 3170-3176.
- [2] Eichengreen, B. (2011). *Exorbitant privilege: The rise and fall of the dollar and the future of the international monetary system*. Oxford University Press.
- [3] Wilson, P. (2015). Monetary policy and financial sector development. *The Singapore Economic Review*, 60(03), 1550031.
- [4] Uddin, I., Siddiq, Q. U., Tabash, M. I., Qayyum, A., & Asadullah, M. (2021). Forecasting Malaysian ringgit against US dollar; individual models vs combined models. *Academy of Accounting and Financial Studies Journal*, 25(6), pp. 1-10.
- [5] Satrio, C. B. A., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, 179, pp. 524-532.
- [6] Cecen, A. A., & Erkal, C. (1996). Distinguishing between stochastic and deterministic behavior in foreign exchange rate returns: *Further evidence*. *Economics Letters*, 51(3), pp.323-329.
- [7] Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning: data mining, inference, and prediction (2nd ed.)*. Springer.

- [8] Shiao, Y. C., Chakraborty, G., Chen, S. F., Li, L. H., & Chen, R. C. (2019). Modeling and prediction of time-series-a case study with forex data. *2019 IEEE 10th international conference on awareness science and technology (ICAST)* (pp. 1-5).
- [9] Ashfaq, T., & Javaid, N. (2019). Short-term electricity load and price forecasting using enhanced KNN. *International Conference on Frontiers of Information Technology (FIT)*, pp. 266-2665. IEEE.
- [10] Ahmad, B., & Zakria, M. (2022). Comparative Study of Box-Jenkins ARIMA and KNN Algorithm for Stock Price Prediction in Pakistan. *Journal of Social Science and Humanities*, 30(1).
- [11] Zhang, D., Wang, X., Gao, L., & Gong, Y. (2021). Predict and analyze exchange rate fluctuations accordingly based on quantile regression model and K-nearest neighbor. In *Journal of Physics: Conference Series (Vol. 1813, No. 1)*, p. 012016. IOP Publishing.
- [12] Kramer, O. (2013). K-nearest neighbors. *Dimensionality reduction with unsupervised nearest neighbors*, pp. 13-23.
- [13] Gupta, T. K., & Raza, K. (2019). Optimization of ANN architecture: a review on nature-inspired techniques. *Machine learning in bio-signal analysis and diagnostic imaging*, pp. 159-182.
- [14] LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning. nature*, 521(7553), pp. 436-444.
- [15] Islam, M. S., Hossain, E., Rahman, A., Hossain, M. S., & Andersson, K. (2020). A review on recent advancements in forex currency prediction. *Algorithms*, 13(8), pp. 186.
- [16] Shahvaroughi Farahani, M., & Razavi Hajiagha, S. H. (2021). Forecasting stock price using integrated artificial neural network and metaheuristic algorithms compared to time series models. *Soft computing*, 25(13), pp. 8483-8513.
- [17] Ayitey Junior, M., Appiahene, P., Appiah, O., & Bombie, C. N. (2023). Forex market forecasting using machine learning: Systematic Literature Review and meta-analysis. *Journal of Big Data*, 10(1), pp. 9.
- [18] Sirignano, J., & Cont, R. (2019). Universal features of price formation in financial markets: perspectives from deep learning. *Quantitative Finance*, 19(9), pp. 1449-1459.
- [19] Sinsomboonthong, S. (2022). Performance comparison of new adjusted min-max with decimal scaling and statistical column normalization methods for artificial neural network classification. *International Journal of Mathematics and Mathematical Sciences*, 2022.
- [20] Yang, M., Wang, H., Jiang, Y., Lu, X., Xu, Z., & Sun, G. (2020). Geca proposed ensemble-knn method for improved monthly runoff forecasting. *Water Resources Management*, 34(2), pp. 849-863.
- [21] Dohnal, V., Kuca, K., & Jun, D. (2005). What are artificial neural networks and what they can do?. *Biomedical papers-palacky University in*, 149(2), pp. 221.
- [22] Tsai, C. P., & Lee, T. L. (1999). Back-propagation neural network in tidal-level forecasting. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 125(4), pp. 195-202.
- [23] Zhang, E., Hou, L., Shen, C., Shi, Y., & Zhang, Y. (2015). Sound quality prediction of vehicle interior noise and mathematical modeling using a back propagation neural network (BPNN) based on particle swarm optimization (PSO). *Measurement Science and Technology*, 27(1), p. 015801.