VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i1.5349

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

TICKER SYMBOL IDENTIFICATION WITH CIMA ON NON-STATIONARY STOCK PRICE DATASET

Aji Gautama Putrada[1]; Maman Abdurohman[2*]; Doan Perdana[3]; Hilal Hudan Nuha[4]

Advanced and Creative Networks Research Center[1,3]
School of Computing[2,4]
Telkom University[1,2,3,4]
www.telkomuniversity.ac.id[1,2,3,4]
ajigps@telkomuniversity.ac.id[1], abdurohman@telkomuniversity.ac.id[2*],
doanperdana@telkomuniversity.ac.id[3], hilalnuha@telkomuniversity.ac.id[4]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

The creation is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License.

Abstract— *Ticker symbol identification based on stock price data in investor decisions has been proven to be pivotal. Though research exists on stock price forecasting, ticker symbol identification is still a research opportunity. Meanwhile, some temporal-sequential classification methods are available, such as classification-integrated moving average (CIMA) and recurrent neural network (RNN)-based deep learning such as long short-term memory (LSTM), and gated recurrent unit (GRU). Our research aim is to prove that CIMA can perform ticker symbol identification on non-stationary stock price datasets. This research collects ten most well-known stock price dataset from Kaggle and performs pre-processing. Then it designs CIMA with non-stationary data and the benchmark deep learning methods. Both methods are optimized with hyperparameter tuning and model selection between adaptive boosting (AdaBoost) and legacy k-nearest neighbors (KNN). The test results show five non-stationary features in the stock price dataset must go through a differentiation process. Then, AdaBoost has an accuracy of 0.9967 ± 0.001, while KNN has an accuracy of 0.9971 ± 0.001, with no significant difference based on t-test. Meanwhile, AdaBoost has a significantly smaller model size and testing and prediction time than KNN. In benchmarking, CIMA+AdaBoost is superior to the three other methods for accuracy, precision, recall, and f1-score, all of which have a value of 0.996. Our research contribution is ticker symbol identification based on stock price using CIMA on multiple-class sequential classification with non-stationary data. For future research, we advice to perform this method on other stock price data.*

Keywords: *classification-integrated moving average, non-stationary data, stock price, temporal-sequential classification, ticker symbol.*

Intisari— *Identifikasi simbol ticker berdasarkan data harga saham dalam pengambilan keputusan investor terbukti penting. Meskipun terdapat penelitian mengenai perkiraan harga saham, identifikasi simbol ticker masih merupakan peluang penelitian. Sementara itu, beberapa metode klasifikasi sekuensial temporal telah tersedia, seperti classification-integrated moving average (CIMA) dan deep learning berbasis recurrent neural network (RNN) seperti long short-term memory (LSTM), dan gated recurrent unit (GRU). Tujuan dari penelitian kami adalah untuk membuktikan bahwa CIMA dapat mengidentifikasi simbol ticker dalam dataset harga saham non-stasioner. Penelitian ini mengumpulkan sepuluh kumpulan data harga saham paling terkenal dari Kaggle dan memprosesnya terlebih dahulu. Kemudian merancang CIMA dengan data non stasioner dan metode deep learning benchmark. Kedua metode dioptimalkan dengan penyetelan hyperparameter dan pemilihan model antara peningkatan adaptif (AdaBoost) dan k-nearest neighbours (KNN) lama. Hasil pengujian menunjukkan lima fitur non stasioner pada dataset harga saham harus melalui proses diferensiasi. Kemudian AdaBoost mempunyai akurasi sebesar 0,9967 ± 0,001, sedangkan KNN mempunyai akurasi sebesar 0,9971 ± 0,001, tidak terdapat perbedaan yang signifikan berdasarkan uji t. Sedangkan AdaBoost memiliki ukuran model serta waktu pengujian dan prediksi yang jauh lebih kecil dibandingkan KNN. Dalam benchmarking, CIMA+AdaBoost lebih unggul dari ketiga metode lainnya dalam hal*

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5349

akurasi, presisi, recall dan f1-score yang semuanya memiliki nilai 0,996. Kontribusi penelitian kami adalah identifikasi simbol ticker berdasarkan harga saham menggunakan CIMA dalam klasifikasi berurutan beberapa kelas dengan data non stasioner. Bagi penelitian selanjutnya, kami menyarankan untuk melakukan metode ini pada data harga saham lainnya.

Kata Kunci*: klasifikasi-rata-rata bergerak terintegrasi, data non-stasioner, harga saham, klasifikasi temporal-sekuensial, simbol ticker.*

INTRODUCTION

Stock price predictions are important because investors have various needs. Singh *et al.* [1] mentioned that stock price predictions is needed for investment decisions. Chen *et al.* [2] mentioned that stock price prediction is used for risk management in harsh environments with high uncertainty. Shah *et al.* [3] said that stock price prediction can stabilize the financial system. Several studies have applied stock price predictions using advanced methods such as deep learning and obtained satisfactory performance results [4]. On the other hand, in making investor decisions, several studies also reveal the importance of identifying stock price ticker symbols [5].

Several studies have carried out ticker symbol identification, where this research generally uses sentiment analysis. An example is research from Khandelwal *et al.* [6], who extracted ticker symbols from Twitter using sentiment analysis and then identified which ones were most discussed on social media. On the other hand, there is a research opportunity to identify ticker symbols from stock price data. The method used is signal identification, which detects and classifies signals in a dataset [7]. Long short-term memory (LSTM) is a deep learning method that is a state-of-the-art method in signal detection [8].

Furthermore, a research previously created a method called classification-integrated moving average (CIMA) which can classify controlling light status based on passive infrared (PIR) sensor signals [9]. The identified weakness is that this method has not been tested in other case studies beside smart lighting data. There is a research opportunity to apply CIMA to identify ticker symbols in stock prices and benchmark them using state-of-the-art signal identification methods. The stock price data is known as non-stationary data [10], where it is important to prove that the integral components in CIMA significantly influence ticker symbol identification.

Our research aim is to prove that CIMA can perform ticker symbol identification on non-stationary stock price datasets. This research collects the most well-known stock price dataset from Kaggle and then performs pre-processing on the data. Then it designs CIMA with non-stationary data and designs the benchmark recurrent neural network (RNN)-based deep learning methods, namely LSTM, RNN, and gated recurrent unit (GRU). Both methods are optimized with hyperparameter tuning and model selection between adaptive boosting (AdaBoost) and k-nearest neighbor (KNN). Finally, the performance of the models are compared.

To the best of our knowledge, there has never been any research that discusses ticker symbol identification using stock price data. The following is a list of our contributions:

1. A stock price dataset consisting of six features: “Open,” “High,” “Low,” “Close,” “Adj Close,” and “Volume,” from ten different ticker symbols: AMZN, AAPL, GOOGL, HYMTF, KO, META, NLFX, SSNLF, SBUX, and TSLA.
2. A ticker symbol identification method using CIMA based on stock price data.
3. An application of CIMA to multiple class problems with non-stationary temporal sequential data performs better than RNN-based deep learning methods.

The remainder of this paper is the following systematics: Section 2 discusses our research design. Section 3 shows our research results and discusses the contributions to related research. Finally, Section 4 presents the conclusions of our study.

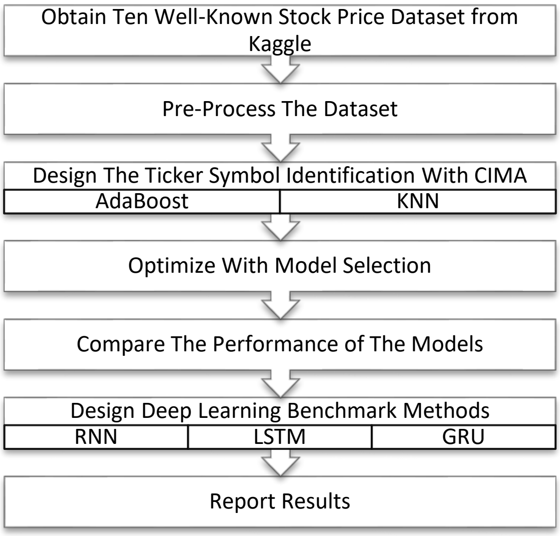
MATERIALS AND METHODS

A research method to achieve our goals is proposed. The most well-known stock price dataset from Kaggle is collected and then pre-processing on the data is performed. The CIMA with non-stationary data and the benchmark RNN-based deep learning methods are designed, namely LSTM, RNN, and GRU. Both methods are optimized with hyperparameter tuning and model selection between AdaBoost and KNN. Finally, the performance of the models are compared. Source: (Research Results, 2024)

Figure 1 presents our research method in the form of a block diagram.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i1.5349



Source: (Research Results, 2024)
Figure 1. Our proposed research method.

1. Stock Price Datasets

A stock price dataset collects information about stock prices in the stock market [11]. Stock price datasets usually have seven columns: ' Date,' 'Open,' 'High,' 'Low,' 'Close,' 'Adj Close,' and 'Volume' [12]. 'Date' is the date the data was taken. 'Open' and 'Close' are the share prices at the beginning and end of that date. 'High' and 'Low' are the highest and lowest stock prices on that date. 'Adj Close' means adjusted close, where the value is the 'Close' price adjusted for other circumstances such as cooperative action. Finally, 'Volume' shows the total value of transactions on that day.

Several studies have used several stock price datasets from Kaggle. Xu *et al.* [13] used the Netflix stock price dataset for research on forecasting using linear regression, decision trees, and gradient boosting. Omoware *et al.* [14] used two datasets, namely Google and Amazon stock price datasets, for LSTM predictions. Chatterjee et al. *et al.* [15] used forecasting to predict the stock price of Hyundai and Samsung using the LSTM method. Thuan *et al.* [16] used Meta and Tesla stock price datasets to predict opening and closing values using methods including autoregressive integrated moving average (ARIMA), support vector regression (SVR), linear regression (LR), and GRU.
Furthermore, Nabi *et al.* [17] classified stock price differences between Starbucks and Apple prices using several feature engineering and machine learning classification techniques. Lastly, Wang *et al.* [18] said that RNN is the most suitable method for carrying out regression on coca-coal stock price prediction. Ten stock price datasets used by previous research are downloaded from Kaggle. Figure 2 shows the first “Open”, “High”, “Low”, and “Close” in the dataset downloaded to provide an illustration of the data.

Source: (Research Results, 2024)
Figure 2. The first 50 stock price data from Kaggle containing “Open,” “High,” “Low,” and “Close” features.

2. Ticker Symbols and Pre-Processing

Every company selling stock on the stock market has a ticker symbol. Ticker symbols are several letters that represent the company [19]. For example, the ticker symbol for Starbucks is SBUX. Several studies show that creative and easy-to-remember ticker symbols influence stock traders' desire to buy that stock [20].

Identifying ticker symbols from stock price data sets may seem niche, but their real-world impact can be significant in several aspects. We have reviewed thoroughly previous literatures on what the impact can bring specifically in the realms of economic information science:

1. Financial and Investment Analysis [21]: Accurate identification ensures that the right stocks are analyzed, resulting in sound investment strategies and decision-making.
2. Market Research and Trading Strategy [22]: Researchers, traders, and financial institutions rely on accurately identifying ticker symbols for market research.
3. Data Integrity and Cleansing [23]: Identifying ticker symbols contributes to the audit party's contribution to data integrity and cleansing in financial data sets.
4. Algorithms and Trading Automation [24]: Accurate identification of ticker symbols is becoming an important part of an automated trading strategy system.
5. Risk Assessment and Portfolio Management [25]: Accurately identified ticker symbols impact a better-managed portfolio.
6. Financial Education and Literacy [26]: Ticker symbol identification can empower individuals to understand financial markets and make informed investment decisions.

Pre-processing of the data is carried out. First, the data are labeled based on each ticker symbol. Then, all data items that have NaN values are dropped. Feature correlation analysis with Pearson correlation is performed[27]. Feature

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5349

selection with a mean decrease in impurity (MDI) is then applied to the dataset [28].

3. CIMA with Differentiation for Non-Stationary Data

In time series analysis, non-stationary data must be differentiated to become stationary [29]. This is because carrying out time-series analysis on non-stationary data is challenging. The augmented Dicky-Fuller (ADF) test is a test to check whether time-series data is stationary or non-stationary [30]. This test can detect whether there is a unit root in a time-series data. If data has a unit root, then the data is non-stationary. The first step of the ADF test is to create the following regression equation:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 y_{t-1} + \cdots + \delta_{p-1} y_{t-p+1} + \varepsilon_t \quad (1)$$

where Δy_t is the first-order difference of the time-series data, α is a constant, t is the lag, β is for the possible coefficient of the lag, γ is the coefficient of the lag, y is the value of the time-series data, δ is coefficient of the lag difference, p is the number of lags, and ε is the error.

Furthermore, from the formula above, the t-statistic from the ADF test has the following formula:

$$t - statistic = \frac{\hat{\gamma}}{Standard\ Error\ of\ \hat{\gamma}} \quad (2)$$

where $\hat{\gamma}$ is the estimated value of γ. From the standardization of the t-statistic, the *p-value* is obtained. If *p-value* $< \alpha$, then H_0 is rejected, and the time-series data is stationary, i.e., there is no rooting value in the dataset. If it is proven that a feature in a dataset is non-stationary, then the feature is differentiated. The formula for the differential for each data item y'_n is as follows:

$$y'_n = \Delta y_n = y_n - y_{n-1}, n \in 2,3,4, \dots, N \quad (2)$$

where N is the dataset size.

The CIMA windowing process is then carried out. CIMA is a classification method for time-series data, where moving averages and differentiation are applied. These methods increase the correlation of time-series data with the classification target labels. The formulas below are used for CIMA windowing:

$$M_j = \left(1 - \frac{j}{J}\right) \times (W - V) + V, j \in J \quad (4)$$

$$\mathrm{MA}_{j,n} = \frac{1}{n} \sum_{i=n-M_j}^{n} y_i, \ \mathrm{n} \in \{J, J+1, J+2, \dots, \mathrm{N}\} \quad (5)$$

where M_j is the size of window j, J is the number of windows, W is the maximum window size, V is the minimum window size, $\mathrm{MA}_{j,n}$, is the moving average of window j and data y_n. J, W, and V are hyperparameters set by the user to obtain optimum CIMA performance.

After completing the CIMA windowing process, feature selection with MDI is performed. The MDI process can prevent overfitting, improve performance, and reduce the model size of the classification model [31]. MDI uses an extra trees classifier to get its impurity value [32]. MDI score is calculated with the following two equations:

$$D_k = \sum_n I_{n-1} - I_n, k \in K \quad (6)$$

$$\mathrm{MDI}_f = \frac{1}{K} \sum_{k=1}^{K} D_k \quad (7)$$

where $f \in F$ and F are the number of features, K is the number of decision trees in the extra tree classifier, D is impurity decrease, I is impurity.

The dataset is divided into train data and test data. Original CIMA trained with KNN is a simple and intuitive machine learning method that can perform regression and classification [33]. KNN makes decisions based on k-nearest data from the data to be classified in feature space [34]. KNN is used because this method performs well compared to other methods such as decision trees and naïve Bayes. However, KNN has several weaknesses, including a large model size and a slow and complex prediction time [35].

KNN can be benchmarked with more sophisticated models, such as ensemble methods. Adaptive boosting (AdaBoost) is an example of an ensemble method, where some of our previous research shows that AdaBoost also has good classification performance [36]. AdaBoost is a boosting type learning ensemble that gives greater weight to data items difficult to classify in its iteration process [37]. The weight w_n^t is given based on an iteration's error value ε_t. The following formulas are involved in the AdaBoost process:

$$\varepsilon_t = \sum_{n=1}^{N} w_n^t \times Indicator(y_i \neq h_t(x_i)) \quad (8)$$

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \quad (9)$$

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

$$w_n^{t+1} = w_n^t \times exp(-\alpha_t \times y_i \times h_t(x_i)) \quad (10)$$

where y_i is the actual label value, $h_t(x_i)$ is the predicted label value, $Indicator(y_i \neq h_t(x_i))$ has a value of 1 if a data item is classified incorrectly. A value of 0 if the data item is classified correctly, and α_t is the importance weight.

Machine learning methods that involve sequential data usually go through pre-processing, whereas pre-processing involves methods such as data transformation or feature extraction [38]. Data transformation methods can involve techniques such as windowing, while feature extraction can involve statistical feature extraction [39]. However, using the classification method on sequential data directly without data transformation can show the ability of the classification method to handle sequential data directly. AdaBoost and KNN are used directly on sequential data to observe and compare their performance with the performance of the windowing process.

CIMA is a sequential classification method. CIMA in ticker symbol identification is compared with state-of-the-art methods in sequential classification, namely from the RNN category [40]. RNN is a class of neural networks that effectively captures the sequential properties of data because its cell structure has memory [41]. Some methods under the RNN family are simple RNN, LSTM, and GRU. LSTM is an RNN variation created to overcome the vanishing gradient problem in RNN [42]. Vanishing gradients appear when neural networks try to learn long-term dependencies. GRU is a variation of LSTM that uses a smaller number of cells so that, with capabilities similar to LSTM, GRU has more efficient complexity [43].

RNN-based deep learning has a distinctive layer architecture. Table 1 explains each layer. The input layer in RNN has a unique shape because, before training, the data goes through a transformation, namely windowing with a certain time step (T). The T value and the feature size F influence the input layer size [36]. The next layer is the recurrent layer, whose contents are recurrent cells. The recurrent cell can be an LSTM, RNN, or GRU. The number of cells is obtained through optimization, which is represented with the notation R. Flattening layers have the function of changing the shape of the previous layer to one dimension. This layer facilitates compatibility between feature learning and classification layers [44]. The last layer is the output layer, where SoftMax activation is used. The size of the SoftMax activation output layer is according to the number of labels. The notation L is given to the output size, where L = 10. The SoftMax function is useful for multi-class classification. Its task is to transform an output label into a probability whose sum is 1 [45].

Table 1. The deep learning architecture for RNN, LSTM, and GRU.

No	Layer Name	Activation	Layer Size
1	Input Layer		$[T, F]$
2	Recurrent Layer	Tanh	$R \times T \times F$
3	Flattening Layer	–	$T \times F$
4	Dropout Layer	–	–
5	Output Layer	SoftMax	L

Source: (Research Results, 2024)

Designing neural network architectures such as LSTM can be complex due to the many hyperparameters involved [46]. The number of LSTM units affects the architecture's ability to learn sequential patterns in the data. The number of epochs affects the LSTM model's ability to correct weights to improve model performance in classification. The optimum batch size makes training convergence occur at the right time and makes training time more efficient. The best learning rate guides the model towards convergence with optimum time. Finally, a large training data size can increase model performance and lengthen the training process time.

RESULTS AND DISCUSSION

1. Results

Ten stock price datasets are downloaded from Kaggle, one each from the companies Amazon, Apple, Coca-Cola, Google, Hyundai, Meta, Netflix, Samsung, Starbucks, and Tesla. All sampling rates are one day. All datasets have seven features: “Date,” "Open,” “High,” “Low,” “Close,” “Adj. Close,” and “Volume.” The last thousand data from each dataset is used. This is to avoid imbalance conditions in identification training. Each company has a ticker symbol: AMZN, AAPL, GOOGL, HYMTF, KO, META, NLFX, SSNLF, SBUX, and TSLA. Table 2 shows the enumeration results that are provided for each dataset.

Table 2. The new stock price dataset containing ten companies’ stock price and stock price symbol.

No	Company Name	Dataset Size	Ticker Symbol	Label
1	Hyundai	1,959	HYMTF	0
2	Samsung	1,228	SSNLF	1
3	Apple	10,468	AAPL	2
4	Amazon	6,155	AMZN	3
5	Google	4,858	GOOGL	4
6	Netflix	1,009	NFLX	5
7	Starbucks	7,919	SBUX	6
8	Tesla	1,692	TSLA	7
9	Meta	2,906	META	8
10	Coca-Cola	15,589	KO	9

Source: (Research Results, 2024)

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

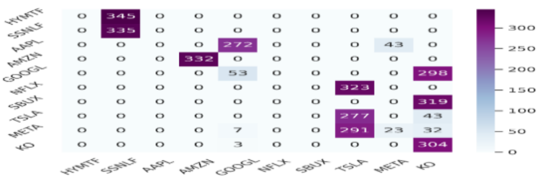
VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5349

The next step sees us perform feature analysis and feature extraction on the dataset. The Pearson correlation coefficient (PCC) and MDI is used on six numerical features in the dataset, where the mean of the MDI score is the threshold for selection. In PCC analysis, all features except volume have the same PCC. Apart from that, all features negatively correlate with the ticker symbol. MDI Score provides more varied results than the PCC score. The mean of the MDI score is 0.167, so the features "Open," "High," and "Low" do not pass feature selection. Table 3 shows the results of our feature analysis and feature selection using the PCC and MDI scores.

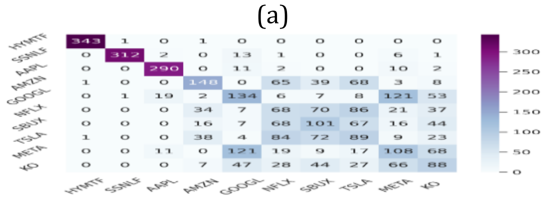
Table 3. Feature analysis and feature selection using PCC score and MDI score.

No	Feature Name	PCC Score	MDI Score	Selected
1	Open	-0,64	0.149	FALSE
2	High	-0,64	0.137	FALSE
3	Low	-0,64	0.121	FALSE
4	Close	-0,64	0.181	TRUE
5	Adj. Close	-0,64	0.181	TRUE
6	Volume	-0,18	0.231	TRUE

Source: (Research Results, 2024)



(a)



(b)

Source: (Research Results, 2024)

Figure 3. Confusion matrix of classification methods without CIMA windowing using (a) AdaBoost (b) KNN

The following test shows how AdaBoost and KNN perform if CIMA windowing was not performed beforehand. A confusion matrix and test size of 0.33 is used to analyze the performance of the two models. **Error! Reference source not found.** shows the confusion matrix of the two models on testing data. AdaBoost has an above-average performance for predicting SSNLF, AMZN, TSLA, and KO. Meanwhile, KNN has an above-average performance in predicting HYMTF, SSNLF, and AAPL. However, both models have below-average performance for predicting other ticker symbols. The accuracy of AdaBoost and KNN are 0.40 and 0.51, respectively.

Next, it is analyzed whether the features used for training are stationary or non-stationary. The ADF test is utilized for this test. The test results influence whether the feature undergoes differentiation or not. Table 4 shows the ADF test results for each feature. ADF test provides p-values above α for "Open," "High," "Low," "Close," and "Adj. Close". This means that five non-stationary features must go through a differentiation process. The "Volume" feature is stationary and has a p-value of 0.002.

Table 4. The ADF test result on each feature.

No	Feature Name	Statistics	P-Value	Conclusion
1	Open	-2.024	0.276	Non-Stationary
2	High	-2,017	0.279	Non-Stationary
3	Low	-2.090	0.248	Non-Stationary
4	Close	-2,019	0.278	Non-Stationary
5	Adj. Close	-2,101	0.244	Non-Stationary
6	Volume	-3,858	0.002	Stationary

Source: (Research Results, 2024)

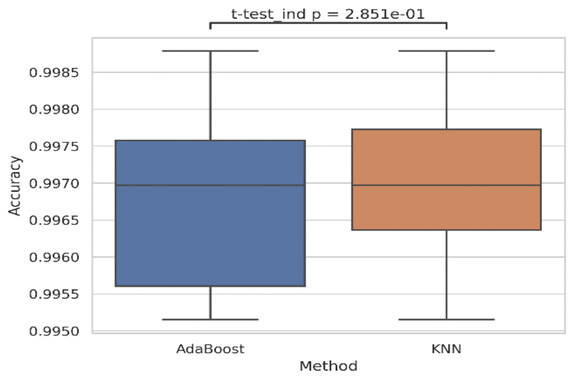
Formula (3) is applied to differentiate the five non-stationary features above so that they all become stationary. These five new features, plus the stationary "Volume" feature, are made into a new dataset. CIMA windowing is then applied to the new dataset. The optimum hyperparameters for ticker symbol identification are $J = 5$, $W = 500$, and $V = 0$. With $J = 5$, CIMA windowing produces 36 new features for ticker symbol identification. Feature selection is performed again for these features using MDI. The results from MDI leave six features, namely the moving average results with $M = 400$ on the differentiated "Close" feature, then the moving average results with $M = 100, 200, 300, 400$, and 500 on the "Volume" feature.

At this stage, 5,000 data items are available. The next step is to do a train test split. The data is divided into training and testing data with a testing size 0.33. Stratifying is applied, namely maintaining the proportion of each label when dividing the data into training and testing data. This is to prevent imbalance from occurring. The result of this step is 3,350 data for training and 1,650 data for testing. There are 335 labels for training data and 165 for testing data. There is no imbalance in the dataset. AdaBoost and KNN training has been conducted twenty times. This is to observe the variability of the two models. The boxplot of accuracy is used to compare the performance of the two models. The t-test is maintained to see the significance of differences in the performance of the two models. Source: (Research Results, 2024)

Figure shows the comparison results of AdaBoost and KNN accuracy for ticker symbol identification. AdaBoost has an accuracy of 0.9967 ± 0.001, while KNN has an accuracy of 0.9971 ± 0.001.

BY NC

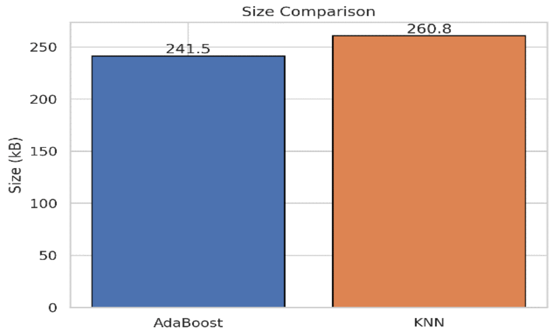
Applying the t-test to both sets of accuracy values gives a p-value of 0.285, above α. These results show no significant difference in the performance of AdaBoost and KNN using CIMA windowing.



Source: (Research Results, 2024)
Figure 4. Accuracy comparison of AdaBoost and KNN with t-test to measure significance in difference on ticker symbol identification with CIMA windowing.

Determining the more optimal model between the two models can use other metrics besides accuracy. The size model for the second metric is used. Source: (Research Results, 2024)

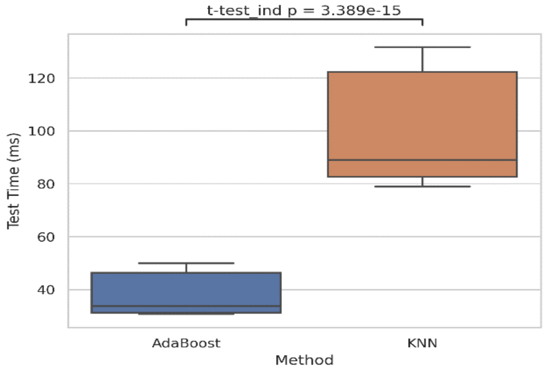
Figure shows the model size comparison between AdaBoost and KNN for ticker symbol identification using CIMA windowing. AdaBoost has a smaller model size than KNN. The AdaBoost model size is 241.5 kB, while the KNN model size is 260.8 kB.



Source: (Research Results, 2024)
Figure 5. Model size comparison between AdaBoost and KNN for ticker symbol identification using CIMA windowing.

Finally, testing time is used as an additional metric to see which model is more optimal for ticker symbol identification. Source: (Research Results, 2024)

Figure shows the comparison results. AdaBoost has a testing time of 37.5 ± 7.3 ms, while KNN has a testing time of 100.4 ± 20.3 ms. The t-test applied to both value distributions have a p-value <0.01 and is smaller than α. AdaBoost has a significantly smaller testing time and prediction time than KNN.



Source: (Research Results, 2024)
Figure 6. Testing time comparison between AdaBoost and KNN with t-test to measure significance in difference on ticker symbol identification with CIMA windowing.

At this stage, the confusion matrix from CIMA with AdaBoost and KNN is analyzed. **Error! Reference source not found.** shows the second image of the confusion matrix. CIMA with AdaBoost has seven false values, while CIMA with KNN has fewer false values, namely four. In CIMA with AdaBoost, the ticker symbol classification with the best recall value belongs to HYMTF, AAPL, SBUX, TSLA, META, and KO. Meanwhile, the CIMA classes with KNN with the best recall are SSNLF, AAPL, AMZN, NLFX, TSLA, and KO. The two confusion matrices show that CIMA windowing can improve the performance of AdaBoost and KNN in ticker symbol identification.

Finally, the architecture of the RNN, LSTM, and GRU models is designed to compare with CIMA+AdaBoost. An exhaustive hyperparameter tuning is performed to get the best performance from the three models. First, it is noticed that increasing the batch size value will also increase the plateau of accuracy in the learning curve. Then it is observed that epochs below 240 and above this value reduce effectiveness. Among stochastic gradient descent (SGD), root mean squared propagation (RMSProp), adaptive movement estimation (Adam), adaptive gradient algorithm (Adagrad), and Adamax, the optimizer that provides the most optimal learning curve is Adamax.

Furthermore, five different learning rates for the optimizer are also varied and the optimum learning rate is 10^{-3}. It is observed that increasing the number of cells in the recurrent layer (R) can improve the model's performance in distinguishing patterns in signals, where the optimum value is 16. Finally, varying the time step in transforming the

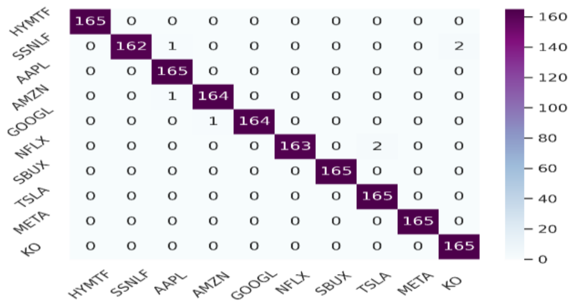
VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5349

input data is attempted. Between the range of 100 to 500, the most optimal value is 200. Table 5 summarizes our optimization.

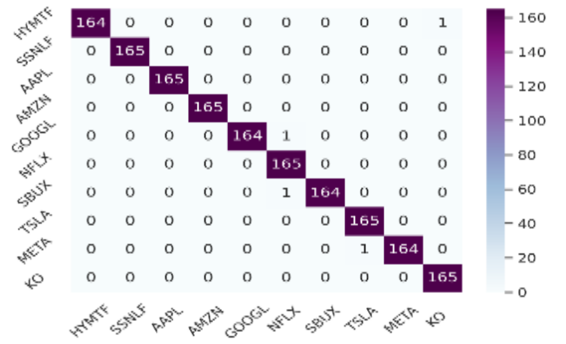
Table 5. Hyperparameter tuning results on RNN-based deep learning models.

No	Hyper-parameter	Tested Values	Optimum Value
1	Batch Size	120, 240, 480, 960, 1920	1920
2	Epoch	30, 60, 120, 240, 480	240
3	Activation Function	SGD, RMSProp, Adam, Adagrad, Adamax	Adamax
4	Learning Rate	$10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$	10^{-3}
5	Recurrent Layer Cells	1, 2, 4, 8, 16	16
6	Time Step	100, 200, 300, 400, 500	200

Source: (Research Results, 2024)



(a)



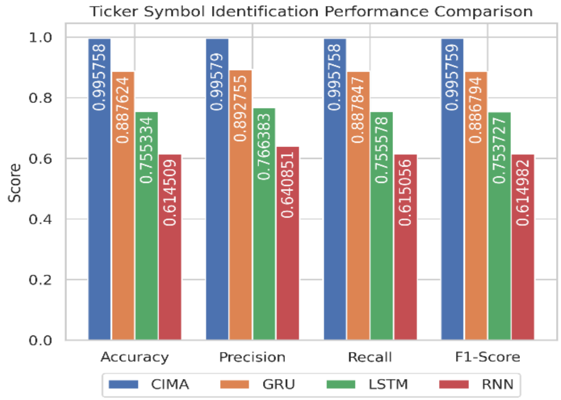
(b)

Source: (Research Results, 2024)

Figure 7. Confusion matrix of classification methods with CIMA windowing using (a) AdaBoost (b) KNN

Source: (Research Results, 2024)

Figure compares the performance of four ticker symbol identification methods: CIMA+AdaBoost, GRU, LSTM, and CNN. Accuracy, precision, recall, and f1-score is used as comparison metrics. CIMA+AdaBoost is superior to the three other methods for accuracy, precision, recall, and f1-score, all of which have a value of 0.996. Among RNN-based deep learning methods, GRU outperforms LSTM and RNN with accuracy, precision, recall, and f1-score values of 0.887, 0.993, 0.887, and 0.887, respectively. LSTM is the deep learning model with the second-best performance, and RNN is the model with the worst performance.



Source: (Research Results, 2024)

Figure 8. Performance comparison of four ticker symbol identification methods.

2. Discussion

Several state-of-the-art studies have used several different stock price datasets. Sharaf et al. [40] researched using a stock price dataset from Amazon. Teng *et al.* [47] used stock prices from 50 different companies, but among them, there were no ticker symbols HYMTF, KO, META, NLFX, SSNLF, SBUX, and TSLA. In addition, the study only used one feature, "Open." Diqi *et al.* [48] used stock prices from PT. Aneka Tambang Tbk. for forecasting and synthetic data creation. This research also did not use the "Adj Close" feature. Mahadik *et al.* [49] used the stock prices of AdaniPorts and Tata Global Beverage Limited. Hossain *et al.* [50] used a stock price dataset of companies with five ticker symbols: BMO, BNS, IMO, SLF, and TRP. In our case, the research contribution is a stock price dataset consisting of six features: "Open," "High," "Low," "Close," "Adj Close," and "Volume," from ten different ticker symbols AMZN, AAPL, GOOGL, HYMTF, KO, META, NLFX, SSNLF, SBUX, and TSLA.

Several previous researchers have studied the ticker symbol research theme. The research results of Prassetio *et al.* [51] showed that changing the ticker symbol can cause the company's stock price to fall dramatically. Nazar *et al.* [52] investigated the relationship between stock value and the likeability and pronunciation of the ticker symbol. The results of this research state that the likeability of the ticker symbol influences stock prices, while the pronunciation does not show a significant effect. Long *et al.* [53] observed the influence of the ticker symbol's congruence with the company name on intangibles in the company. This congruence has an influence on the company, where

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

the ticker symbol is one of the company assets that the company must plan carefully.

Table 6. State-of-the-art research on ticker symbol identification with non-stationary stock price dataset using CIMA

Reference	Stock Price Dataset	Ticker Symbol	CIMA	Non-Stationary Data	Contribution
Sharaf *et al.* [47]	✓	✗	✗	✓	The use of Amazon stock price dataset.
Teng *et al.* [54]	✓	✗	✗	✓	The use of "Open" feature from stock price dataset of 50 ticker symbols, excluding HYMTF, KO, META, NLFX, SSNLF, SBUX, dan TSLA.
Diqi *et al.* [48]	✓	✗	✗	✓	The use of PT. Aneka Tambang Tbk. stock price dataset excluding "Adj Close" feature.
Mahadik *et al.* [49]	✓	✗	✗	✓	The use of AdaniPorts and Tata Global Beverage Limited stock price dataset.
Hossain *et al.* [50]	✓	✗	✗	✓	The use of stock price dataset from BMO, BNS, IMO, SLF, dan TRP ticker symbols.
Prasetio *et al.* [51]	✗	✓	✗	✗	Ticker symbol modification significance on stock price.
Nazar *et al.* [52]	✗	✓	✗	✗	Ticker symbol likeability significance on stock price.
Long *et al.* [53]	✗	✓	✗	✗	Ticker symbol congruence with company name significance on stock price.
Li *et al.* [55]	✗	✓	✗	✗	Ticker symbol similarity correlation with stock price movement similarity.
Li *et al.* [56]	✗	✓	✗	✗	Ticker symbol pronounceability significance on stock price.
Putrada *et al.* [57]	✗	✗	✓	✗	CIMA on binary sequential classification problem with stationary data.
Putrada *et al.* [9]	✗	✗	✓	✗	Nested Markov chain for synthetic data using CIMA.
Proposed Method	✓	✓	✓	✓	1. A stock price dataset consisting of six features: "Open," "High," "Low," "Close," "Adj Close," and "Volume," from ten different ticker symbols: AMZN, AAPL, GOOGL, HYMTF, KO, META, NLFX, SSNLF, SBUX, and TSLA. 2. Ticker symbol identification based on stock price. 3. CIMA on multiple-class sequential classification with non-stationary data.

Source: (Research Results, 2024)

Furthermore, Li *et al.* [55] investigated that similar ticker symbols influence the similarity of stock price movements. Then, the name change also affects stock price movements. Finally, this influence is more on personal investors than institutional investors. The research results of Li *et al.* [56] proved that in Tehran, Iran, the pronounceability of a ticker symbol affects its stock price. Our research contribution is a ticker symbol identification method based on stock price data using CIMA.

Our previous research showed that CIMA was able to improve the classification and control performance of smart lighting on binary temporal sequential data by applying multiple moving average windows [57]. Our other research also shows the power of nested Markov chains in creating synthetic data as CIMA training data [9]. Our current research contribution is an application of CIMA to multiple class problems with non-stationary temporal sequential data, which performs better than RNN-based deep learning methods. **Error! Reference source not found.** summarizes our discussion while highlighting the contribution of each state-of-the-art research while highlighting the research contribution in this paper.

Apart from the scientific impact, our research also has a broad social or real-world impact, including financial analysis, trading strategies, auditing, stock market automation, portfolio management, and education. In the future, the application of CIMA can be directed to stationary multiple-class sequential data, such as human

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5349

activity recognition (HAR) data, and also two-dimensional sequential data, such as images.

CONCLUSION

A dataset of 10 ticker symbols from different companies is formed: AMZN, AAPL, GOOGL, HYMTF, KO, META, NLFX, SSNLF, SBUX, and TSLA. The dataset consists of six features, namely "Open," "High," "Low," "Close," "Adj Close," and "Volume." Ticker symbol identification on the dataset is performed using CIMA, which can be implemented on non-stationary data. Two classification methods is used for comparison: AdaBoost and KNN. Finally, CIMA is benchmarked with three RNN-based deep learning methods: RNN, LSTM, and GRU.

Furthermore, the test results show five non-stationary features in the stock price dataset and must go through a differentiation process. Then, AdaBoost has an accuracy of 0.9967 ± 0.001, while KNN has an accuracy of 0.9971 ± 0.001. Applying the t-test to both sets of accuracy values shows no significant difference in the performance of AdaBoost and KNN using CIMA windowing. On the other hand, AdaBoost has a smaller model size than KNN and has a significantly smaller testing time and prediction time than KNN. In benchmarking, CIMA+AdaBoost is superior to the three other methods for accuracy, precision, recall, and f1-score, all of which have a value of 0.996. Our research contribution is ticker symbol identification based on stock price using CIMA on multiple-class sequential classification with non-stationary data.

Through thorough literature study, we have pointed out that our research also has the potential for a broad social or real-world impact, including financial analysis, trading strategies, auditing, stock market automation, portfolio management, and education. In the future, the application of CIMA can be directed to stationary multiple-class sequential data, such as human activity recognition (HAR) data, and two-dimensional sequential data, such as images.

REFERENCE

[1] A. Singh and V. Devaser, “Analysis For Stock Market Investment Decision Using ML Models,” in *2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, IEEE, 2023, pp. 1–6. Accessed: May 21, 2024, doi: 10.1109/ICIPTM57143.2023.10118262.

[2] X. R. Chen, “Stock price prediction using machine learning strategies,” in *2022 International Conference on Company Management, Accounting and Marketing*, pp. 68–73, 2022.

[3] A. Shah, M. Doshi, M. Parekh, N. Deliwala, and P. P. M. Chawan, “Identifying Trades Using Technical Analysis and ML/DL Models,” Apr. 12, 2023. doi: 10.15680/IJIRCCE.2023.1104038.

[4] A. Kumar, S. Hooda, R. Gill, D. Ahlawat, D. Srivastva, and R. Kumar, “Stock Price Prediction Using Machine Learning,” in *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, Greater Noida, India: IEEE, Apr. 2023, pp. 926–932. doi: 10.1109/CISES58720.2023.10183515.

[5] J. W. Luo, “A Study on Stock Graph Recognition Based on Wavelet Denoising and DTW Algorithm,” *Math. Probl. Eng.*, vol. 2021, pp. 1–15, Jan. 2021, doi: 10.1155/2021/6641749.

[6] V. Khandelwal, H. Varshney, and G. Munjal, “Sentiment Analysis based Stock Price Prediction using Machine Learning,” in *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, IEEE, 2024, pp. 182–187. Accessed: Jul. 22, 2024, doi: 10.1109/InCACCT61598.2024.10551038.

[7] L. Wood, K. Anderson, P. Gerstoft, R. Bell, R. Subbaraman, and D. Bharadia, “Deep Learning Object Detection Approaches to Signal Identification,” Nov. 01, 2022, *arXiv*: arXiv:2210.16173. doi: 10.48550/arXiv.2210.16173.

[8] V. Clerico, J. González-López, G. Agam, and J. Grajal, “LSTM Framework for Classification of Radar and Communications Signals,” in *2023 IEEE Radar Conference (RadarConf23)*, IEEE, 2023, pp. 1–6. Accessed: Dec. 26, 2023, doi: 10.1109/RadarConf2351548.2023.10149618.

[9] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, “Synthetic Data with Nested Markov Chain for CIMA-Based Smart Lighting Control Deployment Simulation,” in *2023 11th International Conference on Information and Communication Technology (ICoICT)*, IEEE, 2023, pp. 148–153. Accessed: Dec. 06, 2023, doi: 10.1109/ICoICT58202.2023.10262430.

[10] D. Batabyal, D. Bandopadhyay, B. Sadhukhan, N. Das, and S. Mukherjee, “Exploring stationarity and fractality in stock market time-series,” in *2023 international conference on intelligent systems, advanced computing and communication (ISACC)*, IEEE, 2023, pp.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

1–6. Accessed: Jul. 22, 2024, doi: 10.1109/ISACC56298.2023.10084056.

[11] S. F. Pane, A. G. Putrada, N. Alamsyah, and M. N. Fauzan, "The Influence of The COVID-19 Pandemics in Indonesia On Predicting Economic Sectors," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, IEEE, 2022, pp. 1–6. Accessed: Dec. 26, 2023, doi: 10.1109/ICIC56845.2022.10006897.

[12] 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," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–24, 2020.

[13] X. Xu, "Research on the Stock Price Forecasting of Netflix Based on Linear Regression, Decision Tree, and Gradient Boosting Models," in *2022 2nd International Conference on Business Administration and Data Science (BADS 2022)*, Atlantis Press, pp. 1232–1242, 2022, doi: 10.2991/978-94-6463-102-9_127.

[14] J. M. Omoware, O. J. Abiodun, and A. I. Wreford, "Predicting Stock Series of Amazon and Google using Long Short-Term Memory (LSTM)," *Asian Res. J. Curr. Sci.*, vol. 5, no.1, pp. 205–217, 2023.

[15] I. Chatterjee, J. Gwan, Y. J. Kim, M. S. Lee, and M. Cho, "An NLP and LSTM based stock prediction and recommender system for KOSDAQ and KOSPI," in *Intelligent Human Computer Interaction: 12th International Conference, IHCI 2020, Daegu, South Korea, November 24–26, 2020, Proceedings, Part I 12*, Springer, 2021, pp. 403–413. doi: 10.1007/978-3-030-68449-5_40.

[16] N. D. Thuan, N. M. Nhut, N. T. V. Huong, and D. V. P. Uyen, "Forecasting the Opening and Closing Price Trends of Stock Using Hybrid Models and Artificial Intelligence Algorithm," in *Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications*, vol. 1688, T. K. Dang, J. Küng, and T. M. Chung, Eds., in Communications in Computer and Information Science, vol. 1688. , Singapore: Springer Nature Singapore, pp. 532–546, 2022, doi: 10.1007/978-981-19-8069-5_36.

[17] R. M. Nabi, S. A. Saeed, and A. M. Abdi, "Feature Engineering for Stock Price Prediction," *Int J Adv Sci Technol*, vol. 29, no. 12s, pp. 2486–2496, 2020.

[18] H. Wang, "Stock Price Prediction Based on Machine Learning and Deep Learning Methods," in *Proceedings of the 2nd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2023)*, Atlantis Press, pp. 1087–1098. 2023, doi: 10.2991/978-94-6463-198-2_112.

[19] D. Ress, "Stock Market Expansion and the Ticker's Siren Song," in *The Kansas Blue Sky Act of 1911*, Cham: Springer Nature Switzerland, 2023, pp. 125–137. doi: 10.1007/978-3-031-43831-8_9.

[20] P. Goyal and M. Goyal, "Impact of corporate name change on stock prices: evidence from BSE," *Gyan Manag. J.*, vol. 17, no. 1, pp. 85–94, 2023.

[21] V. K. Mishra, R. Binyala, P. Sharma, and S. Singh, "Predictive Analysis of Stock Prices Through Scikit-Learn: Machine Learning in Python," in *Mathematics and Computer Science Volume 1*, 1st ed., S. Ghosh, M. Niranjanamurthy, K. Deyasi, B. B. Mallik, and S. Das, Eds., Wiley, pp. 397–404, 2023, doi: 10.1002/9781119879831.ch20.

[22] R. Rodríguez-Aguilar and J.-A. Marmolejo-Saucedo, "Identification of Trading Strategies Using Markov Chains and Statistical Learning Tools," in *Trends in Data Engineering Methods for Intelligent Systems*, vol. 76, J. Hemanth, T. Yigit, B. Patrut, and A. Angelopoulou, Eds., in Lecture Notes on Data Engineering and Communications Technologies, vol. 76. , Cham: Springer International Publishing, pp. 732–741, 2021, doi: 10.1007/978-3-030-79357-9_67.

[23] J. L. Buchanan and M. D. Piercey, "How simple changes to language and tick marks can curtail the ghost ticking of audit procedures," *Audit. J. Pract. Theory*, vol. 40, no. 4, pp. 53–77, 2021.

[24] M. R. Roostaee and A. A. Abin, "Forecasting financial signal for automated trading: an interpretable approach," *Expert Syst. Appl.*, vol. 211, p. 118570, 2023.

[25] L.-J. Liu, W.-K. Shen, and J.-M. Zhu, "Research on risk identification system based on random forest algorithm-high-order moment model," *Complexity*, vol. 2021, pp. 1–10, 2021.

[26] J. Shang, S. Cui, C. Ke, Z. Xia, S. Fu, and D. Liu, "Symbol rate identification using asynchronous delayed sampling," in *CLEO: QELS_Fundamental Science*, Optica Publishing Group, 2015, pp. JTh2A-55. Accessed: May 21, 2024, doi: 10.1364/CLEO_AT.2015.JTh2A.55.

[27] N. Sinha, R. K. Tripathy, and A. Das, "ECG beat classification based on discriminative multilevel feature analysis and deep learning

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5349

approach," *Biomed. Signal Process. Control*, vol. 78, p. 103943, 2022.

[28] A. G. Putrada and N. G. Ramadhan, "MDIASE-Autoencoder: A Novel Anomaly Detection Method for Increasing The Performance of Credit Card Fraud Detection Models," in *2023 29th International Conference on Telecommunications (ICT)*, IEEE, pp. 1–6, 2023, doi: 10.1109/ICT60153.2023.10374051.

[29] R. Baidya and S.-W. Lee, "Addressing the Non-Stationarity and Complexity of Time Series Data for Long-Term Forecasts," *Appl. Sci.*, vol. 14, no. 11, p. 4436, 2024.

[30] C. Fowler, X. Cai, J. T. Baker, J.-P. Onnela, and L. Valeri, "Testing unit root non-stationarity in the presence of missing data in univariate time series of mobile health studies," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 73, no. 3, pp. 755–773, 2024.

[31] A. G. Putrada, N. Alamsyah, S. F. Pane, and M. N. Fauzan, "Feature Importance on Text Analysis for a Novel Indonesian Movie Recommender System," in *2023 11th International Conference on Information and Communication Technology (ICoICT)*, IEEE, pp. 34–39, 2023, doi: 10.1109/ICoICT58202.2023.10262504

[32] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "EdgeSL: Edge-Computing Architecture on Smart Lighting Control With Distilled KNN for Optimum Processing Time," *IEEE Access*, vol. 11, pp. 64697–64712, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3288425.

[33] I. D. Oktaviani and A. G. Putrada, "KNN imputation to missing values of regression-based rain duration prediction on BMKG data," *J. Infotel*, vol. 14, no. 4, pp. 249–254, 2022.

[34] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "Q8KNN: A Novel 8-Bit KNN Quantization Method for Edge Computing in Smart Lighting Systems with NodeMCU," in *Intelligent Systems and Applications*, vol. 824, K. Arai, Ed., in Lecture Notes in Networks and Systems, vol. 824. , Cham: Springer Nature Switzerland, 2024, pp. 598–615. doi: 10.1007/978-3-031-47715-7_41.

[35] R. I. Fakhruddin, M. Abdurohman, and A. G. Pu-trada, "Improving PIR Sensor Network-Based Activity Recognition with PCA and KNN," in *2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, pp. 138–143, 2021

[36] M. N. Fauzan, A. G. Putrada, N. Alamsyah, and S. F. Pane, "PCA-AdaBoost Method for a Low Bias and Low Dimension Toxic Comment Classification.," in *2022 International Conference on Advanced Creative Networks and Intelligent Systems (ICACNIS)*, IEEE, pp. 1–6, 2022, doi: https://doi.org/10.1109/ICACNIS57039.2022.10055017.

[37] A. N. Iman, A. G. Putrada, S. Prabowo, and D. Perdana, "Peningkatan Kinerja AMG8833 sebagai Thermocam dengan Metode Regresi AdaBoost untuk Pelaksanaan Protokol COVID-19," *J. Elektro Dan Telekomun. Terap. E-J.*, vol. 8, no. 1, pp. 978–985, 2021, doi: https://doi.org/10.25124/jett.v8i1.3894.

[38] B. Pavithra, M. Niranjanamurthy, G. V. Smitha, R. Kiran, and M. Chandrika, "Transactional and Sequential model for pre-processing the textual data," in *2023 4th International conference on intelligent engineering and management (ICIEM)*, IEEE, pp. 1-6, 2023, doi: 10.1109/ICIEM59379.2023.10167357

[39] M. Lupión, J. Medina-Quero, J. F. Sanjuan, and P. M. Ortigosa, "DOLARS, a Distributed On-Line Activity Recognition System by Means of Heterogeneous Sensors in Real-Life Deployments—A Case Study in the Smart Lab of The University of Almería," *Sensors*, vol. 21, no. 2, p. 405, Jan. 2021, doi: 10.3390/s21020405.

[40] L. Zhao, Y. Zuo, and K. Yada, "Sequential classification of customer behavior based on sequence-to-sequence learning with gated-attention neural networks," *Adv. Data Anal. Classif.*, vol. 17, no. 3, pp. 549–581, 2023.

[41] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "Recurrent Neural Network Architectures Comparison in Time-Series Binary Classification on IoT-Based Smart Lighting Control," in *2022 10th International Conference on Information and Communication Technology (ICoICT)*, IEEE, pp. 391–396, 2022.

[42] C. B. Vennerød, A. Kjærran, and E. S. Bugge, "Long Short-term Memory RNN," May 14, 2021, *arXiv*: arXiv:2105.06756. Accessed: Jan. 03, 2024. [Online]. Available: http://arxiv.org/abs/2105.06756

[43] A. G. Putrada, N. Alamsyah, S. F. Pane, and M. N. Fauzan, "GRU-MF: A Novel Appliance Classification Method for Non-Intrusive Load Monitoring Data," in *2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Solo, Indonesia: IEEE,

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

Nov. 2022, pp. 200–205. doi: 10.1109/COMNETSAT56033.2022.9994409.

[44] M. Varlı and H. Yılmaz, "Multiple classification of EEG signals and epileptic seizure diagnosis with combined deep learning," *J. Comput. Sci.*, vol. 67, p. 101943, 2023.

[45] P. Privietha and V. J. Raj, "Hybrid Activation Function in Deep Learning for Gait Analysis," in *2022 International Virtual Conference on Power Engineering Computing and Control: Developments in Electric Vehicles and Energy Sector for Sustainable Future (PECCON)*, Chennai, India: IEEE, May 2022, pp. 1–7. doi: 10.1109/PECCON55017.2022.9851128.

[46] A. A. Bataineh and D. Kaur, "Immunocomputing-Based Approach for Optimizing the Topologies of LSTM Networks," *IEEE Access*, vol. 9, pp. 78993–79004, 2021, doi: 10.1109/ACCESS.2021.3084131.

[47] M. Sharaf, E. E.-D. Hemdan, A. El-Sayed, and N. A. El-Bahnasawy, "StockPred: a framework for stock Price prediction," *Multimed. Tools Appl.*, vol. 80, pp. 17923–17954, 2021.

[48] M. Diqi, M. E. Hiswati, and A. S. Nur, "StockGAN: robust stock price prediction using GAN algorithm," *Int. J. Inf. Technol.*, vol. 14, no. 5, pp. 2309–2315, 2022, doi: 10.1007/s41870-022-00929-6.

[49] A. Mahadik, D. Vaghela, and A. Mhaisgawali, "Stock price prediction using lstm and arima," in *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, IEEE, pp. 1594–1601, 2021, doi: 10.1109/ICESC51422.2021.9532655

[50] M. R. Hossain, M. T. Ismail, and S. A. B. A. Karim, "Improving stock price prediction using combining forecasts methods," *IEEE Access*, vol. 9, pp. 132319–132328, 2021, doi: 10.1109/ACCESS.2021.3114809.

[51] Y. H. Prassetio, D. Setiawan, and M. Mutmainah, "Perubahan Nama Perusahaan Terhadap Tingkat Likuiditas dan Abnormal Return Saham," *J. Akunt. Dan Bisnis*, vol. 22, no. 1, pp. 135–153, 2022, doi: 10.20961/jab.v22i1.730.

[52] M. R. Nazar, "Investor Sentiment Stock Price on Indonesia Stock Exchange," *Int. J. Entrep. Bus. Manag.*, vol. 1, no. 2, pp. 139–145, 2022, doi: 10.54099/ijebm.v1i2.376.

[53] Y. Long, W. Su, and Y. Tan, "Does a Share Name Change Have an Impact on the Pricing Efficiency of the Share?," *SAGE Open*, vol. 13, no. 4, p. 21582440231219071, Oct. 2023, doi: 10.1177/21582440231219071.

[54] X. Teng, T. Wang, X. Zhang, L. Lan, and Z. Luo, "Enhancing stock price trend prediction via a time-sensitive data augmentation method," *Complexity*, vol. 2020, pp. 1–8, 2020, doi: 10.1155/2020/6737951.

[55] G. Li, G. Tang, and L. Liu, "Ticker symbol and comovement: Evidence from a unique dataset in China", Accessed: Jan. 07, 2024. [Online]. Available: https://www.researchgate.net/profile/Guangzi-Li/publication/264881512_Ticker_symbol_and_comovement_Evidence_from_a_unique_dataset_in_China/links/57dfe31d08ae484409238fb1/Ticker-symbol-and-comovement-Evidence-from-a-unique-dataset-in-China.pdf

[56] Y. Li, "Study on the relationship between Word Preference and Stock Returns under the background of intelligent information," in *E3S Web of Conferences*, EDP Sciences, p. 01048, 2021, doi: 10.1051/e3sconf/202127501048

[57] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "CIMA: A Novel Classification-Integrated Moving Average Model for Smart Lighting Intelligent Control Based on Human Presence," *Complexity*, vol. 2022, no. Article ID 4989344, p. 19, 2022, doi: 10.1155/2022/4989344.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri