Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 8 No. 1 (2024) 80 - 87 e-ISSN: 2580-0760

Improving Algorithm Performance using Feature Extraction for Ethereum Forecasting

Indri Tri Julianto[1], Dede Kurniadi[2], Ricky Rohmanto[3], Fathia Alisha Fauzia[4]
[1,2]Department of Computer Science, Institut Teknologi Garut, Indonesia
[3]Department of Informatic Management, Universitas Ma'some, Bandung, Indonesia
[4]Department of Communication and Information, Universitas Garut, Indonesia
[1]indritrijulianto@itg.ac.id, [2]dede.kurniadi@itg.ac.id, [3]rickyrohmanto@masoemuniversity.ac.id, [4]fathiaalisha@uniga.ac.id

Abstract

Ethereum is a cryptocurrency that is now the second most popular digital asset after Bitcoin. High trading volume is the trigger for the popularity of this cryptocurrency. In addition, Ethereum is home to various decentralized applications and acts as a link for Decentralized Finance (DeFi) transactions, Non-Fungible Tokens (NFTs) and the use of smart contracts in the crypto space. This study aims to improve the performance of the forecasting algorithm by using Feature Extraction for Ethereum price forecasting. The algorithms used are Neural Networks, Deep Learning and Support Vector Machines. The research methodology used is Knowledge Discovery in Databases. The dataset used comes from the yahoo.finance.com website regarding Ethereum prices. The research results indicated that the use of Feature Extraction improved the performance of the constructed model. The results show that the Neural Network Algorithm is the best Algorithm compared to Deep Learning and Support Vector Machine. The Root Mean Square Error value for the Neural Network before Feature Selection is 93,248 +/- 168,135 (micro average: 186,580 +/- 0,000) Linear Sampling method and 54,451 +/- 26,771 (micro average: 60,318 +/- 0,000) Shuffled Sampling method. Then after the Feature Selection, the Root Mean Square Error value improved to 38,102 +/- 31,093 (micro average: 48,600 +/- 0,000) using the Shuffled Sampling method. This research bridged the gap by either expanding on prior studies or contributing through the comparison of three forecasting algorithms for cryptocurrency datasets. It also compared two feature extraction algorithms, namely Principal Component Analysis and Independent Component Analysis, and employed the T-Test to conduct a performance difference analysis among algorithm results to determine the best model performance.

Keywords: *algorithms; ethereum; feature extraction; forecasting*

How to Cite: I. Tri Julianto, D. Kurniadi, R. Rohmanto, and F. Alisha Fauzia, "Improving Algorithm Performance using Feature Extraction for Ethereum Forecasting ", J. RESTI (Rekayasa Sist. Teknol. Inf.), vol. 8, no. 1, pp. 80 - 87, Feb. 2024.
DOI: https://doi.org/10.29207/resti.v8i1.5417

1. Introduction

The existence of cryptocurrencies in the last decade The past decade has witnessed significant fluctuations in cryptocurrency exchange rates and market capitalization [1]–[3]. Ethereum (XRP) stands as one of the most prominent blockchain technologies following Bitcoin (BTC) [4]. Ethereum, characterized by its decentralized nature, employs Turing-complete capabilities for constructing and executing smart contracts or distributed systems [5]–[7]. Smart contracts are programs whose code and execution status are stored on the blockchain [8]. Initially, Ethereum was designed solely for use within the Ethereum network, unlike its current role as a payment system [4]. Given its decentralized nature, Ethereum's canonical status is determined by network participants through a consensus mechanism, devoid of a central coordinator. This implies that network participants are obliged to assess each transaction, ranging from the genesis block to network resources, computing power, and substantial storage requirements [8].

Ethereum's ranking is close to the first cryptocurrency, namely Bitcoin. The background of this research is to improve the performance of the best forecasting algorithm using the Feature Extraction method. Feature extraction involves the extraction of characteristics that can depict the necessary information [9]. This operation is performed by utilizing a method to capture attributes from a specific form, and the resulting values are then examined for

Received: 18-01-2023 | Accepted: 13-02-2024 | Published Online: 15-02-2024

subsequent processing [10]. Two types of Feature Extraction Algorithms were used, Principal Component Analysis (PCA) and Independent Component Analysis (ICA). The two algorithms will be compared and sought to determine which one has the best Root Mean Square Error (RMSE) value.

There are several studies have been done before regarding Feature Extraction. The first study employed Feature Extraction using Naive Bayes, where the results indicated that the accuracy of the F1-Score achieved was 0.93, and the Area Under Curve (AUC) value was 0.95 [11]. The second study utilized Feature Extraction on a Support Vector Machine and Gray Level Co-Occurrence Matrix, where the obtained accuracy was 83.2% [12]. The third study employed Feature Extraction on Naive Bayes, yielding an accuracy of 56.33% [13]. The fourth study employed Feature Extraction on a Support Vector Machine and Gray Level Co-Occurrence Matrix, resulting in an accuracy of 90.47% [14]. The fifth study utilized Feature Extraction on the Support Vector Machine and K-Nearest Neighbor algorithms, where the results showed an improvement in the accuracy of both algorithms. In particular, the accuracy of SVM reached the highest value at 88.13% [10]. To make it clearer in understanding previous research, a comparative analysis of the previous technique is presented, as shown in Table 1.

Table 1. Comparative Analysis Of Previous Research [15]

Research	Techniques	Outcome	Evaluation
1	Naïve Bayes	Feature Extraction	F1-Score 0.93 & AUC 0.95
2	SVM & GLCM	Feature Extraction	Accuracy 83.2%
3	Naïve Bayes	Feature Extraction	Accuracy 56.33%
4	GLCM & SVM	Feature Extraction	Accuracy 90.47%
5	SVM & K-NN	Feature Extraction	Accuracy 88.13%
Present	NN, DL & SVM	Feature Extraction	RMSE

(NN) Neural Network, (SVM) Support Vector Machine, (DL) Deep Learning, (GLCM) Grey Level Co-Occurrence Matrix, (K-NN) K-Nearest Neighbours (RMSE) Root Means Square Error

This research filled the gap by either building upon previous studies or contributing by comparing three forecasting algorithms for cryptocurrency datasets. It then compared two feature extraction algorithms, namely Principal Component Analysis and Independent Component Analysis. This research is essential to conduct because previous studies only compared two algorithms using Feature Extraction, while this study applies Feature Extraction to three algorithms, thereby increasing the probability of selecting the best algorithm. The research utilized K-Fold cross-validation with a k-value of 10 for model validation, and the selection of the best sampling technique involved a choice between Linear Sampling and Shuffled Sampling. K-Fold cross-validation was adopted in this study because it aligns with the research objective of enhancing model performance, as it provides a more effective evaluation of machine learning models or algorithms compared to straightforward methods like holdout validation. The evaluation model uses the Root Mean Square Error (RMSE) and a different test is carried out using the T-Test. After obtaining the best algorithm and the best sampling technique, the next step is to search for the best performance improvement using Feature Extraction with the Principal Component Analysis (PCA) and Independent Component Analysis (ICA) Algorithms.

This study aimed to enhance the performance of the forecasting algorithm model, wherein we compared three algorithms: Neural Network, Deep Learning, and Support Vector Machine, utilizing a Feature Extraction Algorithm. Principal Component Analysis and Independent Component Analysis were compared to determine the best Feature Extraction Algorithm, which could then be applied to the modelling of Ethereum cryptocurrency forecasting.

2. Research Methods

This research method uses Knowledge Discovery in Databases, which consists of four main stages: data collection, pre-processing, modelling, and evaluation. The research method chart is presented in graphical form, as shown in Figure 1.

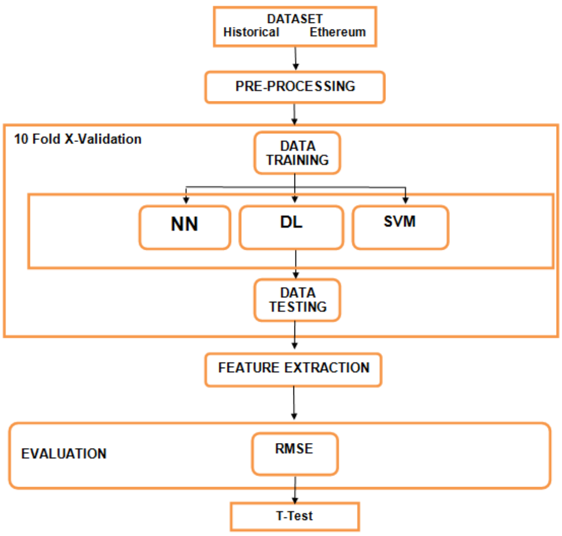

Figure 1. Research Method

The first stage is Dataset. This dataset was compiled from the yahoo.finance.com website [16]. The Ethereum price history matrix in USD is shown in Figure 2.

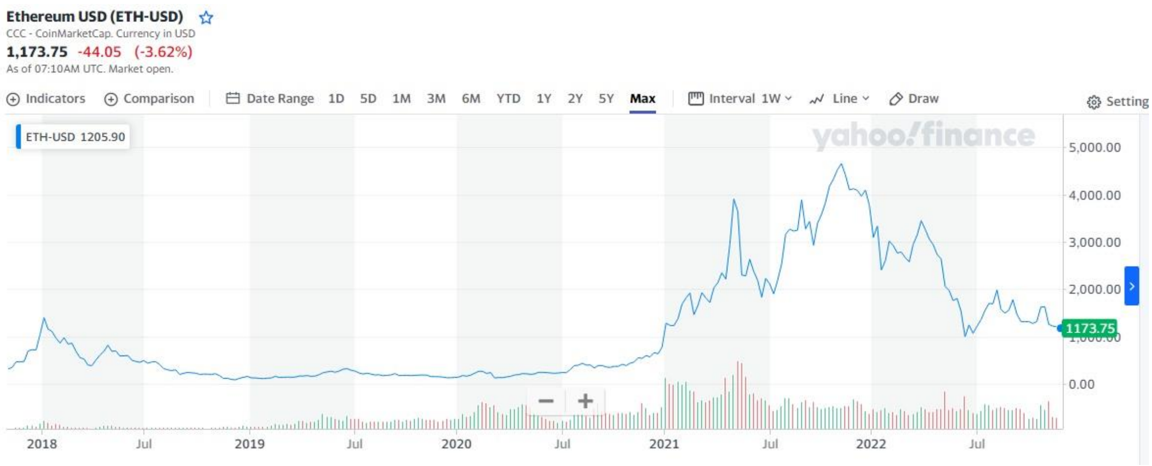

Figure 2. Ethereum Historycal Prices [16]

We can see that in Figure 2, Ethereum prices have increased in mid-2021 and decreased in mid-2022. The dataset collected is population data regarding the history of Ethereum prices in USD from September 2015 to November 2022, where there are 7 attributes, namely: Date = Date (Format Day - Month - Year); Open = Opening Price; High = Highest Price; Low = Lowest Price; Close = Closing Price; Volume = Transaction volume is usually in the number of sheets; Adjusted Close = Closing price adjusted for corporate actions such as rights issue, stock split or stock reverse.

The Second stage is Pre-Processing. This stage will carry out data cleaning and selection of the attributes needed in the Data Mining process [17]. There are 3 methods in Pre-Processing, namely: 1). Data Cleansing, is the process of cleaning data from empty values, inconsistent, empty attributes such as missing values and noisy data; 2). Data Reduction is removing unnecessary attributes; 3). Data Integration, namely merging data into one archive [18].

The Third Stage is Modeling. This Stage will first compare the 3 Algorithms: Neural Network, Deep Learning and Support Vector Machine. The validation model uses K-Fold cross-validation (KCV) where the value of k-10 is compared with the best sampling method between Linear Sampling & Shuffled Sampling. KCV will partition k parts of data and do as many k iterations. Whenever a part of the dataset is selected, the first k – 1 are used as learning data while the rest are used as testing data. This process will be repeated k-times and then the average deviation (error) value of the k-different test results will be calculated [19].

The Fourth Stage is Fetaure Extraction. This stage uses Principal Component Analysis (PCA) [20] and Independent Component Analysis (ICA).

Calculating the Mean ($\bar{X}$) of the data in each dimension using Equation 1.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (1)$$

n is the Number of sample data, and X_i is the sample data.

Calculating the covariance matrix (C_x) using Equation 2.

$$C_x = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})T \quad (2)$$

n is the Number of sample data, X_i is the Sample data, $\bar{X}$ is the Mean.

Calculating the eigenvector (v_m) and eigenvalue(λ_m) of the covariance matrix using Equation 3.

$$C_x v_m = \lambda_m v_m \quad (3)$$

Sort the eigenvalues in descending order. The Principal Component (PC) is a series of eigenvectors according to the order of eigenvalue in stage 3.

ICA is an algorithm that can independently distinguish elements or components of a mixed signal. ICA uses Equation 4[21].

$$X = AS \quad (4)$$

S is a collection of m signal sources, A represents the coupling mechanism source signals, and X represents the mixed signal of the combined result.

The Fifth Stage is Root Mean Squared Error (RMSE). RMSE is a sum of squared errors or differences between actual (actual) value and predicted value, then divide that number by the amount of time and forecasting data then pull out the roots, or you can formulate as Equation 5[22].

$$RMSE = \frac{\sqrt{\sum(Actual - Prediction)^2}}{n} \quad (5)$$

The Sixth Stage is the T-Test. The T-test is a parametric test used to determine whether there is a difference in the means between two related samples [23]. The T-test was employed to examine the differences in RMSE values among algorithms.

3. Results and Discussions

Table 2. Preliminary Data

Date	Close	Open	High	Low	Adj.Close	Volume
9/1/2015	1.35348	1.39274	0.557062	0.738644	0.738644	19920655
10/1/2015	0.734307	1.34505	0.420897	0.916627	0.916627	19815600
11/1/2015	0.920847	1.10642	0.725665	0.873119	0.873119	25989045
12/1/2015	0.878316	1.04537	0.770488	0.933542	0.933542	14436832
1/1/2016	0.933712	2.83699	0.929835	2.30604	2.30604	99427369
2/1/2016	2.31969	6.64741	2.11631	6.33699	6.33699	350388940
3/1/2016	6.31931	15.2571	6.30885	11.4035	11.4035	911491280
4/1/2016	11.4006	11.8504	7.05898	8.81398	8.81398	503439580
5/1/2016	8.77508	15.0482	8.43236	14.0773	14.0773	866285490
6/1/2016	14.1098	21.5227	9.96364	12.4615	12.4615	1179208426
7/1/2016	12.4381	14.8989	9.57626	11.8759	11.8759	851472184
8/1/2016	11.8948	12.4578	8.20098	11.6724	11.6724	507070296
9/1/2015	1.35348	1.39274	0.557062	0.738644	0.738644	19920655
10/1/2015	0.734307	1.34505	0.420897	0.916627	0.916627	19815600

Table 2 shows the initial data that has been collected from the yahoo.finance.com site. The next step is to look for correlations between each attribute in the dataset. The matrix for correlation values between attributes is presented as shown in Table 3.

Then the Correlation Matrix results for the Ethereum dataset are presented as shown in Table 4.

Table 3. Rule Of Thumb About Correlation Coefficient [24]

Coefficient Range	Strange of Association
±0.91to ±1.00	Very Strong
±0.71to ±0.90	High
±0.41to ±0.70	Moderate
±0.21to ±0.40	A small but definite relationship
±0.01to ±0.20	Slight, almost negligible

Table 4. Correlation Matrix

Attribute	Date	Open	High	Low	Close	Adj.Close	Volume
Date	1	?	?	?	?	?	?
Open	?	1	0.983	0.976	0.955	0.955	0.637
High	?	0.983	1	0.978	0.981	0.981	0.703
Low	?	0.976	0.978	1	0.985	0.985	0.631
Close	?	0.955	0.981	0.985	1	1	0.671
Adj.Close	?	0.955	0.955	0.985	1	1	0.671
Volume	?	0.673	0.703	0.631	0.671	0.671	1

Based on Table 4, we can see that the correlation between attributes is at a moderate to very strong association with values for the attributes Open = 1, High = 0.983, Low = 0.976, Close = 0.955, Adj Close = 0.955, and Volume = 0.637. Only the Volume attribute is at a Moderate level.

The rest of the other attributes are at the Very Strong Association level. This shows that the dataset is feasible to be carried out in the following process. The following procedure labels the closed attribute. The results of this process are presented as shown in Table 5.

Table 5. The Dataset Pre-Processing Result

Date	Close	Open	High	Low	Adj.Close	Volume
9/1/2015	1.35348	1.39274	0.557062	0.738644	0.738644	19920655
10/1/2015	0.734307	1.34505	0.420897	0.916627	0.916627	19815600
11/1/2015	0.920847	1.10642	0.725665	0.873119	0.873119	25989045
12/1/2015	0.878316	1.04537	0.770488	0.933542	0.933542	14436832
1/1/2016	0.933712	2.83699	0.929835	2.30604	2.30604	99427369
2/1/2016	2.31969	6.64741	2.11631	6.33699	6.33699	350388940
3/1/2016	6.31931	15.2571	6.30885	11.4035	11.4035	911491280
4/1/2016	11.4006	11.8504	7.05898	8.81398	8.81398	503439580
5/1/2016	8.77508	15.0482	8.43236	14.0773	14.0773	866285490
6/1/2016	14.1098	21.5227	9.96364	12.4615	12.4615	1179208426
7/1/2016	12.4381	14.8989	9.57626	11.8759	11.8759	851472184
8/1/2016	11.8948	12.4578	8.20098	11.6724	11.6724	507070296
9/1/2015	1.35348	1.39274	0.557062	0.738644	0.738644	19920655
10/1/2015	0.734307	1.34505	0.420897	0.916627	0.916627	19815600

In Table 5, it can be observed that the 'close' attribute was utilized as the label. This was done to enable calculations using the Forecasting Model. Forecasting can be executed when the data structure consists of numerical attributes, has numerical labels, and incorporates a time series. Thus, with a data structure like that presented in Table 5, it was ready for modelling. The next stage is to create a model in the Rapidminer Studio application. The model building is presented in the form of an image, as shown in Figure 4.

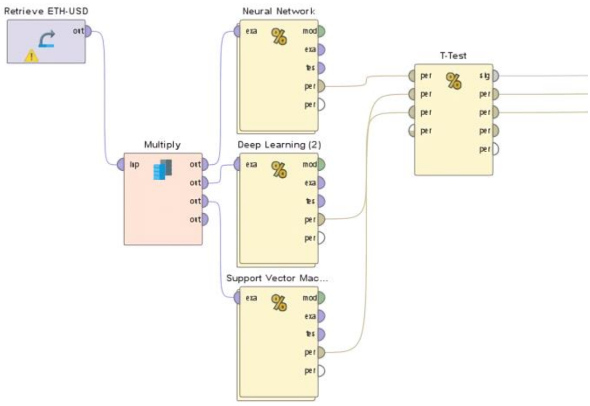

Figure 4. Model Process

The model is built using a dataset that has gone through the Pre-Processing stage and then connected to the Multiply operator because there is more than one Algorithm used. Then the three algorithms, namely Neural Network, Deep Learning, and Support Vector Machine, are connected to the T-Test operator to find out the difference in the performance values of each Algorithm. If the value > 0.050 means there is a significant difference between the algorithms and vice versa. If the value is <0.050, then there is no significant difference, meaning the algorithms are in the same order [19]. The results of the Running model are presented in tabular form, as shown in Table 6, and Figure 5, and the T-Test results are presented in Table 7 and Figure 6.

The analysis of Table 7 and Figure 5 revealed valuable insights into the performance of various algorithms in this study. The data illustrated that the Neural Network stood out as the most effective algorithm, as evidenced by the significantly low Root Mean Square Error (RMSE) value compared to other algorithms. The RMSE value for the Neural Network, utilizing the Shuffled Sampling method, was recorded at 54,451 +/- 26,771 (micro average: 60,318 +/- 0,000), indicating a high precision in its predictive capabilities. Furthermore, when evaluating algorithms based on their error rates, the Support Vector Machine algorithm also exhibited less-than-optimal performance. Specifically, the Support Vector Machine algorithm, using the Shuffled Sampling method, obtained an error rate of 1035.345 +/- 547.894 (micro average: 1166.188 +/- 0,000), which was the lowest among the two other algorithms.

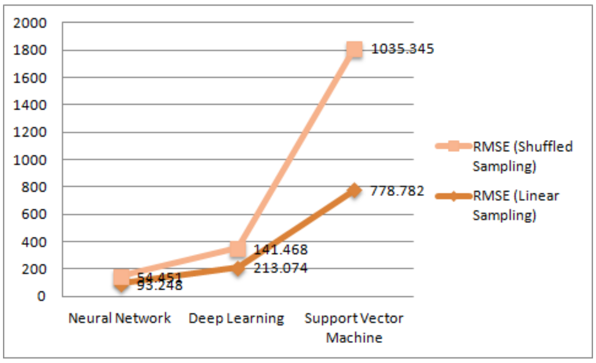

Figure 5. Result RMSE Algorithms

In summary, the detailed examination of Table 7 and Figure 5 underscored the Neural Network as the standout algorithm, excelling in predictive accuracy with the Shuffled Sampling method. Meanwhile, the Support Vector Machine algorithm did not show superior performance compared to the other two algorithms.

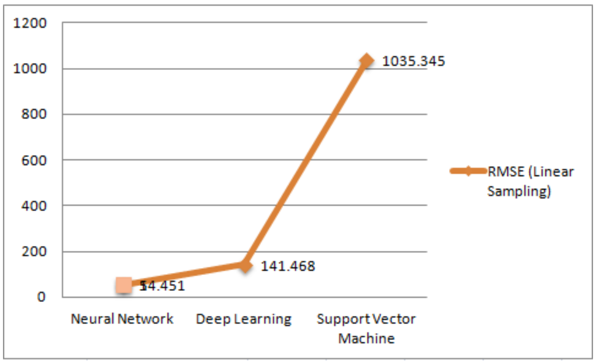

Figure 6. Algorithms Rating Result

Table 6. Performance Model Result Before Feature Extraction

No	Algorithm	RMSE (Linear Sampling)	RMSE (Shuffled Sampling)
1	Neural Network	93.248 +/- 168.135 (micro average: 186.580 +/- 0.000)	54.451 +/- 26.771 (micro average: 60.318 +/- 0.000)
2	Deep Learning	213.074 +/- 292.560 (micro average: 352.192 +/- 0.000)	141.468 +/- 54.832 (micro average: 151.074 +/- 0.000)
3	Support Vector Machine	778.782 +/- 920.856 (micro average: 1174.677 +/- 0.000)	1035.345 +/- 547.894 (micro average: 1166.188 +/- 0.000)

Table 7. Performance Model Ranking Based on T-Test

No	Algorithm	RMSE (Shuffled Sampling)	T-Test
1	Neural Network	54.451 +/- 26.771 (micro average: 60.318 +/- 0.000)	Significant Difference
2	Deep Learning	141.468 +/- 54.832 (micro average: 151.074 +/- 0.000)	Significant Difference
3	Support Vector Machine	1035.345 +/- 547.894 (micro average: 1166.188 +/- 0.000)	Significant Difference

The results of the T-Test show that the Neural Network, Deep Learning, and Support Vector Machine Algorithms have significant differences because they have an alpha value > 0.050, so the Neural Network Algorithm is in the first place. Deep Learning is in second place, and Support Vector Machine is in third place. The next step is to improve the performance of the Neural Network Algorithm by using Feature Extraction (PCA and ICA). The way Feature Extraction works is to perform calculations and comparisons that are used to classify an image based on its histogram characteristics [10]. The process model built on Rapidminer is presented as shown in Figure 7.

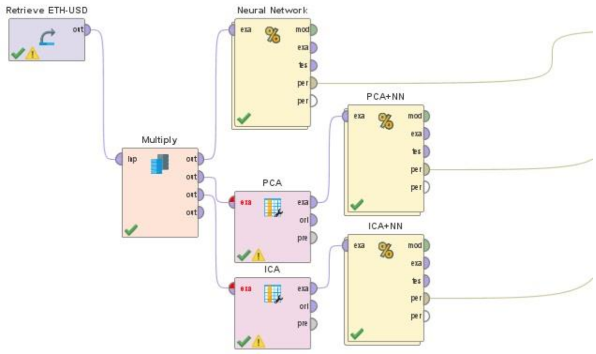

Figure 7. Model Process with Feature Extraction

The model was built to compare the resulting RMSE values between Neural Networks (without feature extraction), Neural Networks with PCA, and Neural Networks with ICA. The sampling method used is Linear Sampling and Shuffled Sampling methods. The K-Fold Cross Validation operator with a value of K=10 is used to display the performance of each Algorithm. Then the PCA and ICA Operators are connected to the K-Fold Cross Validation Operator so that it will immediately see the Root Mean Square Error value when it is run. When using PCA and ICA, it means that the existing attributes in the dataset were merged into new attributes, as seen in Figure 8 for the PCA results.

In Figure 8, we can see that the new attributes have replaced the old ones, where the new attributes are labelled as pc_1, pc_2, up to pc_5. Furthermore, the values in each attribute have also changed after calculations using PCA, where the old attributes were amalgamated into the new ones. To view the dataset resulting from ICA, it is presented as shown in Figure 9.

Row No.	Close	pc_1	pc_2	pc_3	pc_4	pc_5
1	0.739	-2751545203...	314.519	4.920	-16.174	6.166
2	0.917	-2751546254...	314.832	5.443	-16.412	6.251
3	0.873	-2751484519...	314.796	5.376	-15.988	6.272
4	0.934	-2751600041...	314.748	5.465	-15.935	6.262
5	2.306	-2750750136...	313.373	6.168	-17.097	6.361
6	6.337	-2748240520...	309.239	7.741	-18.680	5.926
7	11.403	-2742629497...	300.776	8.172	-20.842	7.547
8	8.814	-2746710014...	299.092	3.565	-16.581	5.655
9	14.077	-2743081555...	297.218	8.819	-18.749	6.239

Figure 8. The new attribute after performing PCA

Row No.	Close	ic_1	ic_2	ic_3	ic_4	ic_5
1	0.739	0.465	-0.231	0.680	-0.186	0.310
2	0.917	0.465	-0.232	0.680	-0.186	0.313
3	0.873	0.463	-0.231	0.681	-0.184	0.311
4	0.934	0.463	-0.231	0.681	-0.184	0.310
5	2.306	0.465	-0.232	0.677	-0.185	0.318
6	6.337	0.470	-0.241	0.668	-0.185	0.325
7	11.403	0.472	-0.223	0.664	-0.184	0.349
8	8.814	0.474	-0.232	0.672	-0.184	0.303
9	14.077	0.470	-0.238	0.665	-0.177	0.329

Figure 9. The new attribute after performing ICA

Similar to PCA, when using ICA, the old attributes were amalgamated into new attributes, labelled as ic_1, ic_2, up to ic_5. The values for each attribute also changed by the calculations performed by the ICA algorithm. After modelling using both PCA and ICA, the comparative values are presented, as shown in Table 8. The analysis of Table 8 provides a comprehensive view of the performance enhancements achieved through Principal Component Analysis (PCA) and Independent Component Analysis (ICA). The emphasis is on identifying the lowest Root Mean Square Error (RMSE) value, which indicates the most optimal results in terms of predictive accuracy.

Table 8. Performance Model Result After Feature Extraction

Algorithms	RMSE (Linear Sampling)	RMSE (Shuffled Sampling)
Neural Network	93.248 +/- 168.135 (micro average: 186.580 +/- 0.000)	54.451 +/- 26.771 (micro average: 60.318 +/- 0.000)
Neural Network + (PCA)	79.966 +/- 155.523 (micro average: 169.647 +/- 0.000)	44.014 +/- 26.347 (micro average: 51.033 +/- 0.000)
Neural Network + (ICA)	55.285 +/- 104.845 (micro average: 114.972 +/- 0.000)	38.102 +/- 31.093 (micro average: 48.600 +/- 0.000)

The standout observation from Table 8 is the synergy between the Neural Network Algorithm, ICA, and the Shuffled Sampling method. This combination yields the most optimal RMSE value among the alternatives, specifically 38,102 +/- 31,093 (micro average: 48,600 +/- 0,000). This outcome signifies a substantial improvement in predictive accuracy compared to other configurations, underscoring the effectiveness of incorporating both ICA and the Shuffled Sampling method into the Neural Network Algorithm.

The utilization of ICA in conjunction with the Shuffled Sampling method appears to contribute significantly to reducing the overall RMSE, implying a more refined and precise prediction model. This finding suggests that the Neural Network Algorithm benefits from the feature extraction capabilities provided by ICA and the enhanced robustness conferred by the Shuffled Sampling method.

The detailed examination of Table 8 illuminates the superior performance achieved through the strategic integration of PCA, ICA, and the Shuffled Sampling method. Specifically, the combination of the Neural Network Algorithm with ICA and the Shuffled Sampling method emerges as the most effective, resulting in a highly optimized RMSE value and exemplifying the importance of thoughtful algorithm and method selection in enhancing model accuracy.

After obtaining the best model, namely the Neural Network Algorithm with Independent Component Analysis, the next step was to conduct a trial using a new dataset to predict the value of Ethereum cryptocurrencies. The trial results were presented in the form of a graph, as shown in Figure 10, and the price issued by yahoo.finance.com from the end of 2022 to the end of 2023, as shown in Figure 11.

In Figure 10, we can see that the blue line represents the closing value, while the green line represents the predicted closing value. We can see that the actual value and the predicted value do not have a wide range of differences. This indicates that the model used has a low error rate, so the predicted value is not far from the actual value.

Figure 10. Prediction Result

Figure 11. Ethereum Price 2022-2023

In Figure 11, the predicted price is directly proportional to the graph displayed on the yahoo.finance.com website regarding the value of the Ethereum cryptocurrency in the range of the end of 2022 to the end of 2023, where there was an increase at the end of 2023, namely from September to December 2023.

4. Conclusions

In this study, we successfully demonstrated that the use of Feature Extraction could enhance the performance of algorithms. We engaged three algorithms Neural Network, Deep Learning, and Support Vector Machine whose performance was enhanced using PCA and ICA. The outcomes of the performance improvement, employing Feature Extraction, revealed that the most optimal RMSE value for the Neural Network was achieved by utilizing Independent Component Analysis (ICA), with a value of 38,102 +/- 31,093 (micro average: 48,600 +/- 0.000). This indicates that Independent Component Analysis (ICA) can enhance the performance of the process model built in this study. This study has several limitations. First, this study focuses on Algorithm improvement by using Feature Extraction. Future research can use other methods for Algorithm improvement, such as Feature Selection. Second, this study uses three Forecasting Algorithms: Neural Network, Deep Learning, and Support Vector Machine. Future research can compare more Forecasting Algorithms to get more optimal results. Third, this study uses the Ethereum cryptocurrency dataset. Future research can add other cryptocurrencies such as Bitcoin and so on.

References

[1] V. Derbentsev, V. Babenko, K. Khrustalev, H. Obruch, and S. Khrustalova, "Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices," *Int. J. Eng. Trans. A Basics*, vol. 34, no. 1, pp. 140–148, 2021, doi: 10.5829/IJE.2021.34.01A.16.

[2] I. T. Julianto, D. Kurniadi, F. A. Fauziah, and R. Rohmanto, "Improvement of Data Mining Models using Forward Selection and Backward Elimination with Cryptocurrency Datasets," *J. Appl. Intell. Syst.*, vol. 8, no. 1, pp. 100–109, 2023.

[3] R. Setiawan, I. T. Julianto, and F. F. Roji, "Time Series Forecasting of Top 3 Ranking Cryptocurrencies," *J. Appl. Intell. Syst.*, vol. 8, no. 2, pp. 193–205, 2023.

[4] N. A. Hitam, A. R. Ismail, and F. Saeed, "An Optimized Support Vector Machine (SVM) based on Particle Swarm Optimization (PSO) for Cryptocurrency Forecasting," *Procedia Comput. Sci.*, vol. 163, pp. 427–433, 2019, doi: 10.1016/j.procs.2019.12.125.

[5] D. L. Kuo Chuen, L. Guo, and Y. Wang, "Cryptocurrency: A New Investment Opportunity?," *SSRN Electron. J.*, vol. 20, no. 3, pp. 16–40, 2018, doi: 10.2139/ssrn.2994097.

[6] A. Prasetya, F. Ferdiansyah, Y. N. Kunang, E. S. Negara, and W. Chandra, "Sentiment Analisis Terhadap Cryptocurrency Berdasarkan Comment Dan Reply Pada Platform Twitter," *J. Inf. Syst. Informatics*, vol. 3, no. 2, pp. 268–277, 2021, doi: 10.33557/journalisi.v3i2.124.

[7] N. Alnuaimi, A. Almemari, M. Madine, K. Salah, H. Al Breiki, and R. Jayaraman, "NFT Certificates and Proof of Delivery for Fine Jewelry and Gemstones," *IEEE Access*, vol. 10, pp. 101263–101275, 2022, doi: 10.1109/ACCESS.2022.3208698.

[8] S. Paavolainen and C. Carr, "Security Properties of Light Clients on the Ethereum Blockchain," *IEEE Access*, vol. 8, pp. 124339–124358, 2020, doi: 10.1109/ACCESS.2020.3006113.

[9] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 406–414, 2021, doi: 10.30865/mib.v5i2.2835.

[10] Y. Religia, "Feature Extraxtion Untuk Klasifikasi Pengenalan Wajah Menggunakan Support Vector Machine dan K-Nearest Neighbour," *Pelita Teknol. J. Ilm. Inform. Arsit. dan Lingkung.*, vol. 14, no. 2, pp. 85–92, 2019.

[11] A. Priyambodo and Prihatini, "Evaluasi Ekstraksi Fitur Klasifikasi Teks Untuk Peningkatan Akurasi Klasifikasi Menggunakan Naive Bayes," *J. Ilm. Elektron. dan Komput.*, vol. 13, no. 1, pp. 159–175, 2020.

[12] M. Muhathir, M. H. Santoso, and D. A. Larasati, "Wayang Image Classification Using SVM Method and GLCM Feature Extraction," *J. Informatics Telecommun. Eng.*, vol. 4, no. 2, pp. 373–382, 2021, doi: 10.31289/jite.v4i2.4524.

[13] K. Ayuningsih, Y. A. Sari, and P. P. Adikara, "Klasifikasi Citra Makanan Menggunakan HSV Color Moment dan Local Binary Pattern dengan Naïve Bayes Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 4, pp. 3166–3173, 2019.

[14] M. Saenudin, F. Haq, R. I. Adam, I. Engineering, and S. Program, "Classification of Covid-19 Using Feature Extraction GLCM and SVM Algorithm," *J. Mantik*, vol. 5, no. 1, pp. 179–183, 2021.

[15] I. T. Julianto, "Design And Build Virtual Reality Photography Web-Based To Support Tourism," *J. Electr. Electron. Information, Commun. Technol.*, vol. 3, no. 2, p. 58, Oct. 2021, doi: 10.20961/jeeict.3.2.54833.

[16] Yahoo Finance, "Ethereum USD (ETH-USD)," *yahoo.finance.com*, 2022. https://finance.yahoo.com/quote/ETH-USD/history?period1=1438905600&period2=1669680000&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true.

[17] A. D. Savitri, F. A. Bachtiar, and N. Y. Setiawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2957–2966, 2018.

[18] Mikhael, F. Andreas, and U. Enri, "Perbandingan Algoritma Linear Regression, Neural Network, Deep Learning, Dan K-Nearest Neighbor (K-Nn) Untuk Prediksi Harga Bitcoin," *JSI J. Sist. Inf.*, vol. 14, no. 1, pp. 2450–2464, 2022, [Online]. Available: http://ejournal.unsri.ac.id/index.php/jsi/index.

[19] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "Data Mining Algorithm Testing For SAND Metaverse Forecasting," *J. Appl. Intell. Syst.*, vol. 7, no. 3, pp. 259–267, 2022.

[20] D. Hediyati and I. M. Suartana, "Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro," *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 49–54, 2021, doi: 10.26740/jieet.v5n2.p49-54.

[21] A. P. Putra, N. W. Wiantari, P. M. Novita Dewi, and I. D. M. Bayu Atmaja Darmawan, "Independent Component Analysis (ICA) Dan Sparse Component Analysis (SCA) Dalam Pemisahan Vokal Dan Instrumen Pada Seni Geguntangan," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 8, no. 1, p. 105, 2019, doi: 10.24843/jlk.2019.v08.i01.p13.

[22] F. H. Hamdanah and D. Fitrianah, "Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikra, Kecil, dan Menengah," *J. Nas. Pendidik. Tek. Inform.*, vol. 10, no. 1, p. 23, 2021, doi: 10.23887/janapati.v10i1.31035.

[23] M. F. Chania, O. Sara, and I. Sadalia, "Analisis Risk dan Return Investasi pada Ethereum dan Saham LQ45," *Stud. Ilmu Manaj. dan Organ.*, vol. 2, no. 2, pp. 139–150, 2021, doi: 10.35912/simo.v2i2.669.

[24] T. Sellar and A. A. Arulrajah, "The Role of Social Support on Job Burnout in the Apparel Firm," *Int. Bus. Res.*, vol. 12, no. 1, p. 110, 2018, doi: 10.5539/ibr.v12n1p110.

