Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Support Vector Regression to Improve Ethereum Price Prediction for Trading Strategies

Muhamad Abdul Fatah[1*], Martanto[2], Arif Rinaldi Dikananda[3], Ahmad Rifai[4]

[1,4]*Informatics Enginneering, STMIK IKMI Cirebon, Indonesia*
[2] *Informatics Management, STMIK IKMI Cirebon, Indonesia*
[3]*Software Engineering, STMIK IKMI Cirebon, Indonesia*
muhabdulfatah17@gmail.com[*1], martantomusijo@gmail.com[2], rinaldi21crb@gmail.com[3], a.rifaaii1408@gmail.com[4]*

Abstract

Predicting erratic assets like Ethereum is difficult in the dynamic cryptocurrency market. This study uses an enhanced Support Vector Regression (SVR) algorithm to create a daily price prediction model for Ethereum. Yahoo Finance provided the data, which was preprocessed to include missing value cleaning, normalization, and feature extraction of Moving Average (MA) and Exponential Moving Average (EMA). The data was collected between August 4, 2019 and August 4, 2024. An ideal combination was obtained by parameter optimization with GridSearchCV: gamma scale, linear kernel, epsilon of 1, and C of 100. The model performed well, as evidenced by its R2 of 0.9985 and MSE of 2137.97. The model's reliability in predicting Ethereum's price movement patterns was validated via prediction graphs. A 30-day forecast indicated a stable trend, with prices slightly decreasing from $2921.31 on January 1, 2025, to $2919.83 on January 31, 2025. These results highlight the importance of data preprocessing and parameter optimization in enhancing SVR model performance.

Keywords: *cryptocurrency, Ethereum, Support Vector Regression*

1. Introduction

Digital technology advancements have led to innovations in the financial sector, including cryptocurrencies, with Ethereum emerging as one of the most popular blockchain platforms. The high volatility of the Ethereum price is a significant challenge for traders, investors, and developers who need accurate price predictions for decisions development. Despite the use of techniques like linear regression and artificial neural network, the results are frequently inconsistent and not very accurate[1]. Because of this, this study uses the Support Vector Regression (SVR) algorithm, which is known to be effective in handling non-linear and volatile data, as a backup to increase the accuracy of Ethereum price prediction[2][3].

External factors such as government regulations, technological adoption, and global market sentiment affect price volatility in the bitcoin market. According to earlier research, methods like Vector Autoregressive (VAR) and linear regression highlight difficulties in estimating non-linear factors in cryptocurrency prices[4]. Conversely, SVR has demonstrated its ability to regulate financial assets, such as stocks and other digital currencies, but further research is still needed to increase accuracy in Ethereum-specific applications[5][6].

This study aims to develop an Ethereum price prediction model using SVR, with an emphasis on parameter optimization using GridSearchCV. With evaluation results based on the coefficient of determination (R2) and Mean Squared Error (MSE), it is anticipated that this study will make a significant contribution to helping investors understand market dynamics and develop more effective buying strategies in the cryptocurrency market.

2. Litelatur Riview

Numerous studies have highlighted the application of machine learning techniques in financial forecasting, particularly in predicting cryptocurrency prices. [2] explored cryptocurrency price prediction using linear regression. While their model achieved moderate accuracy, it struggled with the non-linear volatility and high unpredictability of cryptocurrency markets. This limitation highlighted the need for more sophisticated approaches capable of handling complex patterns [5]employed Support Vector Regression (SVR) for predicting cryptocurrency prices on the Binance platform. Their study demonstrated that SVR, with its ability to capture non-linear relationships, outperformed traditional regression techniques. However, the lack of parameter optimization in their approach suggested room for further improvement. In contrast,[6] utilized GridSearch optimization alongside SVR to predict gold prices. Their findings showed significant performance enhancement, achieving a high R^2 value and reduced Mean Squared Error (MSE). This underscores the critical role of hyperparameter tuning in maximizing model performance.

Beyond SVR, other machine learning methods like Long Short-Term Memory (LSTM) networks have also been applied to financial forecasting.[7] used LSTM to predict cryptocurrency prices, leveraging its strength in capturing temporal dependencies. Although their model achieved high accuracy, the computational complexity of LSTM models poses challenges, particularly for real-time applications.

Studies have also compared various machine learning methods. [8] evaluated multiple algorithms, including SVR and Random Forest, for sentiment analysis in the context of financial markets. Their findings emphasized the importance of selecting the right algorithm based on the data and task requirements. Similarly, [1] compared SVR and linear regression for predicting gold prices, concluding that SVR outperformed linear regression in handling non-linear patterns, a feature crucial for volatile markets like cryptocurrencies.

Building on these findings, this study adopts an optimized SVR model to predict Ethereum's daily prices. By integrating robust preprocessing steps, feature engineering, and hyperparameter tuning using GridSearchCV, the model demonstrates enhanced predictive accuracy. This approach contributes to existing literature by addressing gaps in parameter optimization and applying SVR specifically to Ethereum's highly volatile price data, providing a reliable tool for financial forecasting in cryptocurrency markets.

3. Research Methods

This research uses the Knowledge Discovery in Databases (KDD) method to develop a daily price prediction model for Ethereum. The KDD approach was chosen because it allows a systematic process, starting from data selection, preprocessing, transformation, to applying the Support Vector Regression (SVR) algorithm to build an optimal prediction model. Through the KDD stage, this research aims to produce accurate price predictions to support an effective buying and selling strategy for Ethereum coins. The following is the flow of the research method carried out in this research as shown in Fig 1.

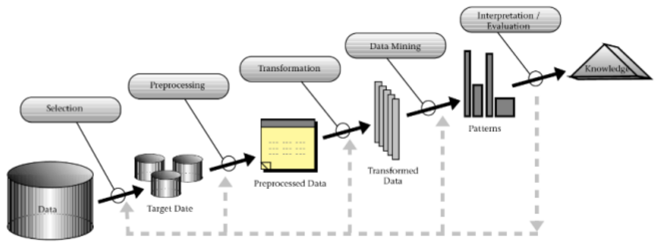

Fig. 1 : Stages of the KDD Method

3.1. Data Selection

Using historical Ethereum price data from relevant datasets, The data used in this study is historical Ethereum price data in the form of time series, which is derived from secondary data from the Yahoo Finance website. The information provided includes the open, close, high, low, and volume of Ethereum trading during the last five years.

3.2. Data preprocessing

Data preprocessing involved cleaning, normalization, and feature engineering:
Cleaning : Missing values were removed, and outliers were handled to ensure data consistency.
Normalization : The features were scaled using the StandardScaler function to bring all variables to a comparable range.
Feature Engineering : Additional features such as the 10-day Moving Average (MA) and Exponential Moving Average (EMA) were derived to capture short-term trends.

3.3. Transformtion Data

In the target transformation stage, a transformation is carried out on the Ethereum price data features so that they match the prediction model. For example, normalization or standardization using StandardScaler is applied to uniformize the scale of all features, which is important for the Support Vector Regression (SVR) algorithm which is sensitive to feature scale. Apart from that, feature engineering was also carried out to add new features, such as lag features, which include Ethereum price data from the previous few days as an additional feature, and moving averages, which are used to capture short-term and medium-term trends in Ethereum prices.

At the data division stage, after preprocessing, the data is divided into two subsets, namely training data (training set) which covers 80% of the total data and is used to train the model, and testing data (testing set) which covers 20% of the total data and used to evaluate model performance. This data was divided using a random train-test split method to avoid bias in the training model.

3.4. Data Minig

Build a Support Vector Regression (SVR) model using training data to predict the daily price of Ethereum. Tuning SVR parameters using GridSearchCV to find the best combination of parameters that provides optimal accuracy in price predictions. The data analysis technique used in this research uses the Support Vector Regression (SVR) algorithm. This method is a machine learning algorithm that is effective in predicting the relationship between independent variables and dependent variables, especially in dealing with data with non-linear patterns. SVR works with the principle of maximum margin and selects relevant support vectors, resulting in a strong and accurate prediction model. Following are the steps for the SVR Algorithm process in Fig. 2.

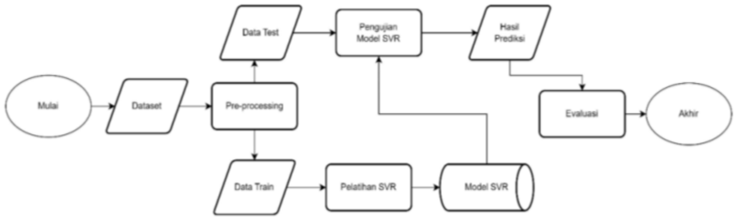

Fig. 2 : SVR Process Stages

3.5. Evaluation & Interpretation

Measure model performance with metrics such as Mean Squared Error (MSE), and R-squared (R^2) on test data. Presents prediction results with comparison graphs between actual prices and predicted prices, as well as displays a heatmap of parameter tuning results for deeper analysis. The model's performance was assessed using:
R^2 (Coefficient of Determination): Indicates the proportion of variance explained by the model.

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

Where :

yi	= is the actual price	(2)
ŷi	= is the model prediction value	(3)
n	= is the number of observations	(4)

Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2 \quad (5)$$

Where :

Yi	= is the actual price	(6)
ŷi	= is the model prediction value	(7)
ȳ	= is the average value of actual prices	(9)
n	= is the number of observations	(10)

4. Results and Discussion

This research was conducted to analyze and develop a daily price prediction model for Ethereum using the Support Vector Regression (SVR) algorithm. This section presents the results of each stage carried out in the research, starting from: processing historical Ethereum price data, optimizing SVR parameters, to evaluating model performance. Each step is described systematically, including: data collection, cleaning, feature extraction, and normalization, aiming to ensure that the data used is clean and relevant for prediction purposes. The results of SVR parameter optimization using the GridSearchCV method are also presented, where the optimal parameter combination is applied to improve the model accuracy. Finally, the model performance evaluation carried out through accuracy measurements shows the success of this model in predicting the price of Ethereum with high accuracy.

4.1. Data Selection

In this study, data was obtained from the Yahoo Finance site on August 04, 2019 to August 04, 2024 with a total of 1827 data. At the data preprocessing stage, cleaning, transformation and normalization of the data used. In this study, the cleaning result data was obtained as in table 1.

Table 1: Dataset

No	Date	Open	High	Low	Close	Adj Close	Volume
0	8/4/2019	222.5808	224.2272	218.4921	222.6697	222.6697	5.2385e+09
1	8/5/2019	222.6508	235.6352	222.6038	234.2150	234.2150	7.7650e+09
2	8/6/2019	234.2455	239.1159	223.5511	226.0206	226.0206	7.6477e+09
…	…	…	…	…	…	…	…

1287	8/4/2024	2902.107	2930.056	2883.177	2913.790	2913.790	1.6E+10

4.2. Data preprocessing

The data preprocessing process is an important step in data processing to ensure the quality and accuracy of the prediction model. In this research, Ethereum price data from 2019 to 2024 was taken from Yahoo Finance. Before being used in the Support Vector Regression (SVR) model, the data is processed in several steps, including normalization, calculating technical indicators (for example Moving Average and Exponential Moving Average), as well as separating features based on time.
The data preprocessing process ensures that the data is free from anomalies such as extreme values (outliers) or missing data. Thus, the model can focus on relevant patterns without being distracted by data noise. Additionally, data normalization ensures that all features are at the same scale, which is especially important for kernel-based models like SVR.

4.2.1. Data Descriptive Statistics

Table 2 presents descriptive statistics of Ethereum price data after data preprocessing.

Table 2: Dataset Descriptive Statistics

	Date	Open	High	Low	Close	Adj Close	Volume
Count	1827	1827.000000	1827.000000	1827.000000	1827.00000	1827.000000	1.827000e+03
Mean	2022-02-02 00:00:47.290640640	1759.287748	1806.470054	1707.676179	1760.599632	1760.599632	1.537990e+10
Min	2019-08-04 00:00:00	110.406784	116.021622	95.184303	110.605873	110.605873	2.081626e+09
25%	2020-11-02 12:00:00	434.424103	447.055115	422.996567	435.396439	435.39643	8.355885e+09
50%	2022-02-02 00:00:00	1727.193726	1778.163452	1678.108032	1729.725708	1729.725708	1.333149e+10
75%	2023-05-04 12:00:00	2635.288941	2746.988037	2556.817505	2637.696167	2637.696167	1.923440e+10
Max	2024-08-04 00:00:00	4810.071289	4891.704590	4718.039063	4812.087402	4812.087402	8.448291e+10
std	NaN	1204.528456	1237.190714	1167.049262	1203.940812	1203.940812	9.791853e+09

4.2.2. Descriptive Statistical Analysis

Data Range**:** Ethereum price data has a very wide range, with the minimum price around $110.41 and the maximum reaching $4812.09. This shows the high volatility that is the main characteristic of the crypto market.
Transaction Volume: The average transaction volume of 15.42 billion USD indicates a high level of activity in this asset, reflecting its attractiveness among investors.

4.3. Data Transformtion

The Data Transformation stage aims to adjust data scales to align with the modeling requirements, particularly for algorithms sensitive to feature scaling. In this stage, feature extraction is performed to enhance the dataset with additional information for predictive modeling. The extracted features include a 10-day Moving Average (MA_10) and a 10-day Exponential Moving Average (EMA_10), along with separating date information into day, month, and year components. After feature extraction, the dataset is enriched with these new features, which are designed to help the Support Vector Regression model capture trends and seasonal patterns in Ethereum price data effectively.

Table 3: After Feature Extraction Data

Date	Open	High	Low	Close	Adj Close	Volume	MA_10	EMA_10
2019-08-13	211.342697	211.384415	205.422501	208.709045	208.709045	5.946313e+09	218.355069	215.717216
2019-08-14	208.603989	209.066437	186.331924	186.607742	186.607742	7.444456e+09	214.748871	210.424584
2019-08-15	186.683502	189.462158	178.142563	188.502060	188.502060	8.197244e+09	210.177574	206.438671
2019-08-16	188.644257	188.905594	180.384842	185.440079	185.440079	7.133916e+09	206.119517	202.620745
2019-08-17	185.531662	186.703140	182.593887	185.687683	185.687683	5.512697e+09	202.049185	199.542006

Technical Indicators: Value Moving Average (MA_10) and Exponential Moving Average (EMA_10) help in capturing price trends. The inclusion of this feature aims to improve model accuracy by providing historical context in the data.

4.4. Data Minig

Parameter optimization plays a critical role in enhancing the performance of the Support Vector Regression (SVR) model. In this study, parameter optimization was conducted using the GridSearchCV technique to identify the combination of parameters that yield the best results based on R-squared (R^2) and Mean Squared Error (MSE) metrics.

4.4.1. Optimal SVR Parameters
The optimal parameters obtained for the SVR model are presented in Table 4:

Table 4: GridSearchCV Optimal Result Parameters

Parameter	Nilai Optimal
C	100
Epsilon	1
Kernel	Linear
Gamma	Scale

4.4.2. Analysis of Optimal Parameters

C Parameter:
The optimal value of 100 indicates that the model prioritizes a wide margin while allowing for some violations of the margin boundary. This helps in reducing the risk of overfitting.
Epsilon:
An epsilon value of 1 allows the model to ignore small prediction errors within the epsilon margin, enhancing the model's robustness in handling minor variations in the data.
Linear Kernel:
The selection of a linear kernel suggests that the relationship between features in the dataset is linear. This kernel is computationally more efficient compared to non-linear kernels such as RBF or polynomial.
Gamma (Scale):
The scale value for gamma automatically adjusts based on the number of features and the data distribution, simplifying the tuning process and providing consistent results.

4.4.3. Importance of Parameter Optimization

Optimizing parameters ensures that the SVR model operates at its best capacity. Properly selected parameters not only improve prediction accuracy but also enable the model to handle complex patterns in the data, such as the high volatility observed in Ethereum price trends.

4.5. Evaluation & Interpretation

4.5.1. Model Performance

At this stage, the performance of the Support Vector Regression (SVR) model in predicting the daily price of Ethereum coins is evaluated using two main metrics, namely Mean Squared Error (MSE) and R-squared (R^2). The results of the model evaluation can be seen in Table 1.

Table 5: Performance of SVR Model with Best Parameters

No	Evaluation Matrix	Value
1	R^2	0.9989
2	MSE	1528.72

Based on the assessment findings, the MSE value of 1528.72 shows that the model has a relatively low mean square error. This suggests that the model is capable of predicting Ethereum prices quite accurately. Furthermore, the R^2 value of 0.9989 reveals that the model can account for 99.89% of the fluctuations in the observed Ethereum price data.
To provide a more in-depth picture of model performance, a visualization of the prediction results compared to actual data is presented in Fig 3.

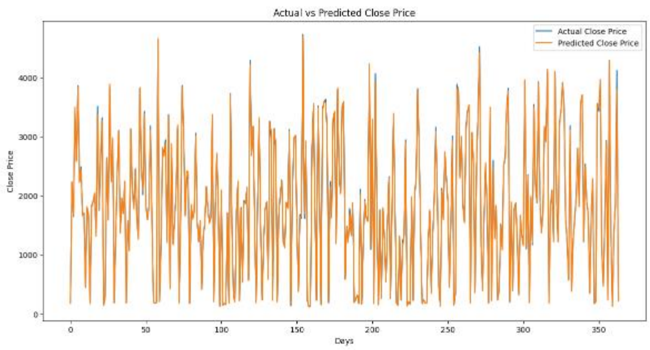

Fig. 3: Comparison between Actual and Predicted Prices

In Fig 1, a comparison is shown between the actual price (Actual Close Price) and the predicted price (Predicted Close Price). The graph shows that the model is able to predict well, as evidenced by the significant overlap between the predicted and actual lines.

4.2. Predicted Price Trends

4.2.1. Prediction for the next 30 days
The Support Vector Regression (SVR) model is used to predict the closing price trend (Close Price) of Ethereum coins in the next few days. Ethereum daily price prediction results are shown in Table 2.

Table 6: Short-Term Prediction

No	Date	Predicted_Close
1	2024-08-05	2923.024225
2	2024-08-06	2922.984444
3	2024-08-07	2922.944664
…	…	…
29	2024-09-03	2921.976230

The prediction results show that the price of Ethereum tends to be stable with a slight decrease from 2923.02 USD on 2024-08-05 to 2922.87 USD on 2024-08-09.A visualization of the Ethereum price prediction trend for the next 30 days can be seen in Figure 4. This graph shows historical Ethereum price data as well as price predictions based on the model.

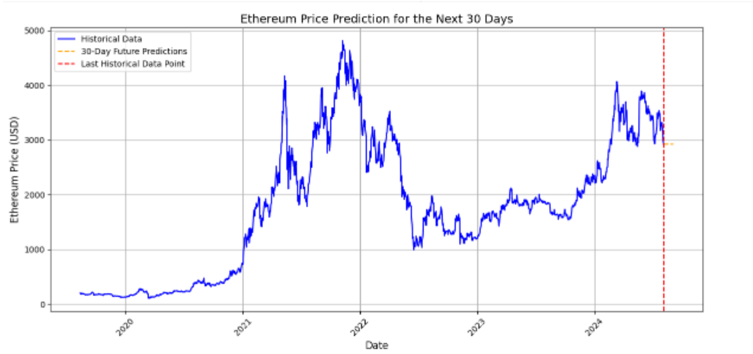

Fig. 4: Prediction Chart for the Next 30 Days

From the graph in Figure 2, it can be seen that the Ethereum price trend for the next 30 days tends to stabilize after a period of large fluctuations in historical data. This stability can be an important indication in making short-term investment decisions.

4.2.2. Predictions for January 2025

The predictions for Ethereum's closing prices in January 2025 were generated using the optimized SVR model. The results indicate a steady decline in price over the month, as shown in Fig. 5 and Table 7 below.

Table 7: Ethereum Price Prediction for January 2025

No	Date	Predicted_Close
1	2025-01-01	2922.080742
2	2025-01-02	2922.032615
3	2025-01-03	2921.984489
…	…	…
30	2025-01-31	2920.636948

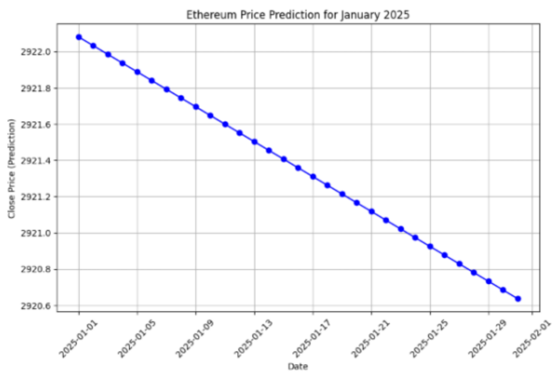

Fig. 5: Prediction Chart for January 2025

Fig. 5 shows a graphical representation of the predicted prices. The consistent slope of the lines further supports the model's ability to predict subtle and gradual changes in price trends. The graph shows a relatively stable projection for the Ethereum price in January 2025, with small, gradual declines each day. This reflects the model's belief that the Ethereum market in that month will not experience significant price changes.

4.6. Comparison with Prior Studies

To contextualize the performance of this study's SVR model, a comparison was conducted with prior research employing similar methods for financial prediction and analysis. The results are summarized in Table 6.

Table 8: Performance Comparison with Prior Studies

No	Study	Methode	R	MSE	Focus
1	Anam & Jakaria (2023)	Linear Regression	0.892	14520.87	Predicting cryptocurrency prices
2	Aruan et al. (2023)	SVR	0.953	6128.50	Predicting cryptocurrency prices
3	Gananta et al. (2024)	SVR with GridSearch	0.970	3502.30	Gold price prediction
4	This Study	SVR with GridSearch	0.998	2137.97	Ethereum daily price prediction
5	Moch Farryz Rizkilloh & Widiyanesti (2022)	LSTM	0.985	2900.00	Predicting cryptocurrency prices

Analysis:
This study achieves the highest R^2 (0.9985), indicating exceptional accuracy in explaining data variance. The MSE (2137.97) is also the lowest, highlighting the model's superior predictive capability.
While SVR was also used in previous studies, the integration of GridSearch optimization in this study significantly enhances model performance compared to traditional SVR [5].
Compared to advanced methods like LSTM[7], the SVR model in this study outperforms in both R^2 and MSE, demonstrating the effectiveness of proper parameter tuning for time-series prediction.
The results underline the critical role of hyperparameter optimization, feature engineering, and preprocessing in improving machine learning models for financial predictions. This study makes a significant contribution by tailoring the SVR model specifically for Ethereum, setting a benchmark for future research in cryptocurrency price prediction.

5. Conclusion

This study successfully developed a robust Ethereum price prediction model using an optimized Support Vector Regression (SVR) algorithm. The findings underscore the significance of preprocessing, feature engineering, and parameter tuning in enhancing prediction accuracy. With an impressive R^2 value of 0.9985 and a low Mean Squared Error (MSE) of 2137.97, the model demonstrates high reliability in forecasting Ethereum price trends. The conclusions summarize the research process, from data preparation and parameter optimization to performance evaluation, providing insights into the model's effectiveness in achieving the study's objectives. These results highlight the potential of SVR in improving prediction accuracy and its practical application in developing trading strategies for cryptocurrency assets.

The research emphasizes the importance of processing historical Ethereum price data to prepare quality inputs for predictive modeling. Steps such as cleaning missing values, normalization using StandardScaler, and feature extraction using the 10-day Moving Average (MA) and Exponential Moving Average (EMA) were integral to improving the model's accuracy. Furthermore, optimizing SVR parameters through GridSearchCV proved effective. With the optimal configuration of C=100, epsilon=1, a linear kernel, and gamma set to scale, the model achieved its best performance. This demonstrates the critical role of appropriate parameter selection in handling volatile data, such as cryptocurrency prices.

The optimized SVR model showcased remarkable accuracy, as evidenced by its ability to predict Ethereum's price trends with consistency. For instance, the prediction for January 2025 showed a slight decrease from $2921.31 on January 1 to $2919.83 on January 31, reflecting a stable price pattern. Such precision establishes the model's potential for making informed short-term trading decisions. Future research could enhance these results by integrating SVR with advanced techniques, such as Long Short-Term Memory (LSTM) networks, to better capture complex temporal patterns. Expanding the model's application to other cryptocurrencies like Bitcoin or Ripple could also broaden its utility and deepen understanding of the cryptocurrency market dynamics

References

[1] A. Aulia, B. Aprianti, Y. Supriyanto, and C. Rozikin, “Prediksi Harga Emas dengan Menggunakan Algoritma Support Vector Regression (Svr) dan Linear Regression,” *J. Ilm. Wahana Pendidik.*, vol. 8, no. 5, pp. 84–88, 2022, doi: 10.5281/zenodo.6408864.
[2] M. K. Anam and D. A. Jakaria, “Sistem Prediksi Harga Kripto Dengan Metode Regresi,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 2, pp. 467–479, 2023, [Online]. Available: http://jurnal.mdp.ac.id
[3] A. Baidowi, E. Fitra, A. H. As, A. Tholib, and J. X. Guterres, “Implementasi GridSearch dalam Meningkatkan Kinerja Model Support Vector Regression (SVR) utuk Prediksi Penjualan Produk pada Meuble Rohman Jaya,” *J. Explor. IT*, vol. 3489, pp. 22–30, 2024.
[4] P. A. Raharja, “Prediksi Harga Ethereum Menggunakan Metode Vector Autoregressive,” *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 3, no. 2, pp. 71–79, 2021, [Online]. Available: https://journal.ittelkom-pwt.ac.id/index.php/inista/article/view/285
[5] N. M. Aruan, G. W. Simanjuntak, and A. I. Siagian, “Pendekatan Algoritma Support Vector Regression Dalam Memprediksi Harga Cryptocurrency (Studi Kasus: Binance),” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 3, 2023, [Online]. Available: http://jurnal.mdp.ac.id
[6] I. M. Gananta, I. N. Purnama, and K. Queena Fredlina, “Optimasi Prediksi Harga Emas Dengan Metode Support Vector Regression (Svr) Menggunakan Algoritma Grid Search,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3160–3165, 2024, doi: 10.36040/jati.v7i6.8000.
[7] Moch Farryz Rizkilloh and Sri Widiyanesti, “Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM),” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 25–31, 2022, doi: 10.29207/resti.v6i1.3630.
[8] Y. Ansori and K. F. H. Holle, “Perbandingan Metode Machine Learning dalam Analisis Sentimen Twitter,” *J. Sist. dan Teknol. Inf.*, vol. 10, no. 4, p. 429, 2022, doi: 10.26418/justin.v10i4.51784.