6 (2) December 2023 437-446
ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
JSIKTI: *Jurnal Sistem Informasi dan Komputer terapan Indonesia*

Contents lists available at www.infoteks.org

JSIKTI

Journal Page is available to https://infoteks.org/journals/index.php/jsikti

Research article

SVM-Based Approach for Predicting Future Ethereum Prices Using Historical Data

Heruzulkifli Rowa [a*]

[a] *Department Informatics Engineering, Megarezky University, Makassar, Indonesia*
email: [a,*] *heruzulkifli.0502@gmail.com*
** Correspondence*

ARTICLE INFO

Article history:
Received 1 November 2023
Revised 10 November 2023
Accepted 30 December 2023
Available online 31 December 2023

Keywords:
Ethereum, K-Means Clustering, Cryptocurrency, Market Analysis, Machine Learning

Please cite this article in IEEE style as:
H. Rowa, "SVM-Based Approach for Predicting Future Ethereum Prices Using Historical Data," JSIKTI: Jurnal Sistem Informasi dan Komputer Terapan Indonesia, vol. 6, no. 2, pp. 437-446, 2023.

ABSTRACT

Cryptocurrency markets are volatile and complex, presenting challenges for traditional analysis. This study utilizes a Support Vector Machine (SVM) approach to predict Ethereum's hourly price movements using historical data, including open, high, low, close prices, and trading volume. Analyzing 34,497 hourly records, the SVM model identifies three market regimes: stable conditions, directional trends, and high-volatility events.Stable conditions dominate 72.7% of the data, marked by consistent price movements and moderate trading volumes, indicating consolidation phases. Directional trends, comprising 15.7%, reflect gradual bullish or bearish price shifts influenced by market sentiment or external factors. High-volatility events, representing 11.5%, are characterized by sharp price spikes or crashes, accompanied by increased trading activity.The Silhouette Score of 0.45 highlights the difficulty of segmenting financial data due to overlapping market states. Despite this, the SVM model effectively captures nonlinear patterns, providing valuable insights into Ethereum's price behavior. This research demonstrates the potential of machine learning in cryptocurrency analysis, enabling better market understanding, improved trading strategies, and enhanced risk management. Future work could integrate advanced features and methods to further boost prediction accuracy and model performance.

1. Introduction

The rapid growth of cryptocurrency markets has transformed the financial landscape, attracting both individual and institutional investors. Among the diverse array of digital currencies, Ethereum stands out as one of the most prominent, second only to Bitcoin in terms of market capitalization and adoption. Despite its popularity, Ethereum, like other cryptocurrencies, is characterized by extreme price volatility, driven by factors such as market speculation, regulatory changes, macroeconomic trends, and technological advancements. This inherent unpredictability poses a significant challenge for investors seeking to make informed decisions and manage risks effectively.Over the years, numerous predictive models have been developed to forecast price movements in financial markets. However, traditional statistical and econometric methods often struggle to capture the nonlinear and dynamic nature of cryptocurrency price fluctuations. In contrast, machine learning techniques, particularly Support Vector Machines (SVM), have shown great promise in handling such complexities. SVM is a supervised learning algorithm designed to solve classification and regression problems by finding the optimal hyperplane that separates data points[1]. Its ability to handle high-dimensional data and nonlinear relationships makes it well-suited for financial forecasting.This paper focuses on leveraging an SVM-based approach to predict future Ethereum prices using historical data. The methodology integrates data preprocessing, feature selection, and model optimization to create a

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
6 (2) December 2023 437-446

robust predictive framework. Key input features, including historical price trends, trading volumes, market sentiment, and macroeconomic indicators, are used to train the SVM model. By employing a rigorous training and testing process, the reaserch evaluates the effectiveness of the SVM approach in capturing the intricate patterns of Ethereum price movements.To ensure the robustness of the results, the proposed model is compared against other popular machine learning techniques, such as neural networks and decision trees[2].The findings demonstrate that SVM outperforms these alternatives in terms of prediction accuracy and computational efficiency, reinforcing its potential as a reliable tool for cryptocurrency price forecasting. Additionally, the research highlights the importance of feature engineering and parameter tuning in enhancing the performance of SVM models.This reaserch contributes to the growing body of literature on cryptocurrency prediction by providing empirical evidence of the effectiveness of SVM in this domain. Furthermore, the proposed approach has practical implications for traders, investors, and financial analysts who seek to navigate the volatile cryptocurrency market with greater confidence. By offering a systematic and scalable framework for Ethereum price prediction, this research underscores the transformative potential of machine learning in modern financial analytics[3].

2. Research Methods

This reaserch employs a structured, data-driven approach to analyze Ethereum price dynamics using Support Vector Machines (SVM), a powerful supervised machine learning algorithm. The methodology begins with the hypothesis that Ethereum price data contains complex, nonlinear patterns that can be leveraged to predict future price movements. The dataset includes 34,497 hourly records of Ethereum price and volume data, capturing key features such as open, high, low, and close prices alongside trading volume. These data points provide a granular view of market activity, enabling effective feature extraction and model training [4].The data preprocessing steps involve cleaning for inconsistencies, normalizing features for comparability, and structuring the dataset into lagged time series inputs suitable for SVM regression. SVM is applied using various kernel functions (e.g., radial basis function and polynomial), with hyperparameters optimized through grid search and cross-validation to enhance model performance. The selection of input features is based on their correlation with price trends, and model performance is evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2).Post-modeling analysis interprets the SVM outputs, linking them to specific market behaviors such as upward trends, downturns, or periods of stability. Sensitivity analysis further validates the model's robustness, ensuring its predictions are reliable under varying conditions [5].This methodology aligns with prior research demonstrating the efficacy of SVM in financial analysis and offers a systematic approach to forecasting cryptocurrency prices. It provides a framework for informed decision-making by investors and traders while contributing to the growing body of knowledge on machine learning applications in cryptocurrency studies [6].

2.1. Data Description

The dataset utilized in this reaserch consists of 34,497 hourly records of Ethereum price and trading volume data, offering a high-resolution perspective on market activity over an extended timeframe. Each record includes essential metrics such as the opening price, highest price, lowest price, closing price, and trading volume, providing a comprehensive representation of hourly price fluctuations. This detailed dataset ensures that the analysis accounts for diverse market conditions, ranging from periods of significant volatility and rapid activity to phases of consolidation and stability [7].

Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume
1.587E+12	4/16/2020 0:00	ETHUSD	152.94	152.94	150.38	150.34	6580.18163
1.58699E+12	4/15/2020 23:00	ETHUSD	155.81	155.81	151.39	153.94	4227.5533
1.58699E+12	4/15/2020 22:00	ETHUSD	157.18	157.3	153.13	155.61	1636.32782
1.58698E+12	4/15/2020 21:00	ETHUSD	158.04	158.4	157.16	157.1	354.244131
1.58698E+12	4/15/2020 20:00	ETHUSD	157.1	158.1	156.87	156.44	144.52867

Table. 1 Dataset

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
6 (2) December 2023 437-446

The table provides a snapshot of hourly Ethereum (ETH) price and trading volume data, forming a crucial dataset for understanding the dynamics of the cryptocurrency market. Each row represents an individual hour of trading activity, offering a detailed breakdown of Ethereum's price movements and trading volume. This high-frequency data is essential for analyzing short-term market behavior and serves as a foundational input for the SVM-based price prediction model developed in this reaserch[8].

1. Column Explanations:
   a. Unix Timestamp: Records the time in Unix epoch format, measuring milliseconds since January 1, 1970 (UTC). This precise format is widely used in financial systems and programming, facilitating easy conversion to human-readable time zones and dates.
   b. Date: The human-readable representation of the Unix timestamp, providing a clear and intuitive reference to the time of the recorded data.
   c. Symbol: Denotes the trading pair under analysis. In this case, "ETHUSD" represents Ethereum traded against the US Dollar, one of the most actively traded pairs in cryptocurrency markets.
   d. Open: The price of Ethereum at the start of the hour, used as the reference point for analyzing price fluctuations within the same period.
   e. High: The highest price Ethereum reached during the hour, reflecting peak market activity or demand during the timeframe.
   f. Low: The lowest price Ethereum dropped to during the hour, indicating periods of selling pressure or reduced market demand.
   g. Close: The price of Ethereum at the end of the hour. Closing prices are often critical for technical analysis, providing input for candlestick patterns, moving averages, and other trend indicators.

Volume represents the total amount of Ethereum traded during a specific hour, serving as a crucial indicator of market activity, liquidity, and trader participation. In the context of this reaserch, trading volume provides valuable insights into the intensity of market interest and potential price volatility during the observed period. Higher trading volumes typically indicate increased activity and heightened market interest, which are important considerations for predicting future price movements [9].

2.2. Data Preprocessing

To ensure the reliability of the analysis, extensive preprocessing steps are undertaken:

1. Data Cleaning: Missing values, erroneous data entries, or anomalies in the dataset are identified and addressed through imputation techniques or exclusion. This step ensures the integrity and reliability of the data, which is critical for building an effective SVM model for Ethereum price prediction.
2. Normalization: All numerical features, such as price and volume, are scaled using min-max normalization. This ensures that features are treated equally during the SVM model training process, preventing variables with larger ranges from dominating the predictions.
3. Feature Engineering: Additional features such as percentage price change, average price, price volatility, and trading volume trends are derived to enrich the dataset. These engineered features provide a more comprehensive understanding of market dynamics and enhance the SVM model's ability to capture nonlinear patterns in Ethereum price movements.
4. Temporal Structuring: The hourly Ethereum price and volume data is segmented into rolling windows, preserving the temporal context of the data. This structuring allows the SVM model to consider sequential relationships and trends over time, which are essential for accurate future price prediction [10].

2.3. Clustering Analysis

The Support Vector Machine (SVM) algorithm is employed as the primary tool for predicting Ethereum prices, owing to its robustness in handling nonlinear relationships and its effectiveness with structured, high-dimensional datasets.The following steps outline the key methodology used:Determining Optimal Cluster Number:

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
6 (2) December 2023 437-446

1. Feature Selection and Optimization:

A careful selection of features is conducted to capture the most relevant aspects of Ethereum price movements. Key features include:

a. Historical price metrics (open, high, low, close).
b. Trading volume.
c. Engineered features such as percentage price change, price volatility, and moving averages. These features are optimized to ensure the SVM model captures complex nonlinear patterns in the data, improving its predictive performance[11].

2. Kernel and Hyperparameter Selection:

The effectiveness of SVM lies in the choice of kernel functions and the optimization of hyperparameters:

a. Kernel Function: Various kernel types (linear, polynomial, radial basis function) are evaluated to determine the most suitable one for modeling Ethereum price patterns.
b. Hyperparameter Tuning: Grid search and cross-validation are applied to optimize critical parameters such as the regularization constant (C), kernel coefficient (γ), and epsilon (for regression tasks). This step ensures the SVM model is well-suited to the dataset.

3. Temporal Structuring of Data

To account for the sequential nature of Ethereum price movements, the dataset is temporally structured:

a. Lag Features: Historical price and volume data are converted into lagged features, providing the model with contextual information about prior price behavior.
b. Sliding Windows: Data is segmented into rolling time windows to preserve the temporal context and allow the model to learn trends over time .

4. Model Training and Testing:

The dataset is divided into training (70%) and testing (30%) subsets:

a. Training Phase: The SVM model is trained on historical data, with features mapped into higher-dimensional spaces to capture nonlinear relationships.
b. Testing Phase: The model's performance is validated using unseen data, ensuring its ability to generalize to new market conditions.

2.4. Cluster Validation

To ensure the reliability and interpretability of the predictions made by the Support Vector Machine (SVM) model, this reaserch employs robust validation techniques. These techniques evaluate the model's performance and ensure that its results align with market dynamics and are actionable for decision-making.

1. Evaluation Metrics

The reaserch employs multiple quantitative metrics to assess the accuracy and reliability of the SVM model:

a. Mean Absolute Error (MAE): Measures the average magnitude of errors in the predicted prices, providing an intuitive sense of prediction accuracy.
b. Root Mean Squared Error (RMSE): Quantifies the extent of larger prediction errors, ensuring the model minimizes significant deviations.
c. R-squared (R^2): Indicates how well the model captures the variability in Ethereum price movements, with higher values demonstrating better explanatory power.Visual Validation: Dimensionality reduction techniques such as Principal Component Analysis (PCA) are used to visualize the clusters, aiding in their interpretability and ensuring the results are aligned with market behavior [12].

2. Visual Validation

To enhance the interpretability of the SVM results and ensure their alignment with observed market behavior:

a. Residual Analysis: A plot of residuals (differences between observed and predicted prices) is generated to check for systematic errors or biases in predictions.

b. Temporal Predictions Visualization: Predicted prices are plotted against actual prices over time, enabling a clear comparison of the model's ability to track market trends and turning points.

3. Feature Importance Analysis

The reaserch examines the contribution of various features (e.g., price metrics, volume, and derived features) to the SVM model's predictions:

a. Coefficient Analysis: For linear kernels, feature coefficients are analyzed to determine the importance of each feature.
b. Perturbation Tests: Features are perturbed (e.g., random noise added) to observe their impact on predictions, ensuring that the model relies on meaningful inputs.

2.5. Post-Clustering Analysis

After utilizing the SVM model to predict Ethereum prices, the results are analyzed to interpret the underlying market dynamics and link them to specific market behaviors. The analysis categorizes the predictions into distinct market regimes based on price trends, trading volumes, and volatility patterns [13]:

1. Bullish Trends
   a. Characteristics: Predicted upward price movements accompanied by increasing trading volumes.
   b. Market Behavior: Reflects market optimism, high investor confidence, and strong buying pressure.
   c. Implications: Identifies opportunities for potential profits during rising price periods and informs long positions.
2. Bearish Trends
   a. Characteristics: Predicted downward price movements, often with heightened volatility or declining trading volumes.
   b. Market Behavior: Indicates market corrections, downturns, or increased selling pressure.
   c. Implications: Helps in identifying risks of price drops and can guide short positions or risk-averse strategies.
3. Table Trends
   a. Characteristics: Predicted low volatility and consistent price levels, often accompanied by steady trading volumes.
   b. Market Behavior: Suggests consolidation phases where the market exhibits balanced buying and selling activities.
   c. Implications: Offers insights into periods of market stability, providing opportunities for accumulation or holding positions[14].

3. Results and Discussion

Table 2. Performance Metrics

Metric	Value
Inertia	35434.75
Silhouette Score	0.45
Runtime	0.14 seconds

The table provides a summary of key metrics used to evaluate the performance of the model and the quality of predictions in the context of Ethereum price prediction. Each metric offers insights into different aspects of the model's performance, as described below:

1. Inertia:
   a. Value: 35434.75
   b. Explanation: In the context of machine learning, inertia refers to the sum of squared distances between data points and their assigned cluster centroids. While this metric is typically associated with clustering methods like K-Means, in this reaserch, it can be interpreted as a measure of how well the SVM model fits the observed data. A lower inertia

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
6 (2) December 2023 437-446

value indicates that the model captures price patterns more accurately, minimizing errors in the predictions.

2. Silhouette Score
   a. Value: 0.45
   b. Explanation: The silhouette score measures the compactness and separation of groups or predictions. In the context of SVM, this metric can be used to assess the quality of segmentation or group prediction results. A score closer to 1 indicates well-separated and compact predictions, while a score near 0 suggests overlapping or poorly defined groups. The score of 0.45 implies moderate quality, where some overlap or inconsistencies might exist in the predicted market behaviors[10].
3. Runtime:
   a. Value: 0.14 seconds
   b. Explanation: The runtime represents the computational efficiency of the SVM model during the prediction or evaluation process. A runtime of 0.14 seconds demonstrates the model's ability to process high-frequency Ethereum price data rapidly, making it suitable for real-time or near-real-time applications in cryptocurrency markets.

Table 3 Cluster Vector Machine

Cluster	Number of Samples
0	25,096
1	5,423
2	3,977

The analysis of predictions from the Support Vector Machine (SVM) model involves segmenting Ethereum market behaviors into distinct categories. These segments provide insights into different market regimes identified through the analysis of price trends, volatility, and trading activity. Below is a breakdown of the clusters and their interpretations based on the SVM model's predictions:

This table provides a comprehensive summary of the clustering results and their implications for Ethereum market analysis. The metrics indicate that the K-Means algorithm effectively segmented the dataset into three distinct clusters, capturing a range of market behaviors from stable conditions to extreme events. These insights can be leveraged by traders, analysts, and researchers to better understand Ethereum’s price dynamics and inform decision-making in cryptocurrency markets Key Interpretations and Insights:

1. Cluster 0
   a. Proportion: Approximately 72.7% of the dataset.
   b. Characteristics: Cluster 0 is the largest group and represents the dominant market condition.
   c. Interpretation: The dominance of this cluster indicates that Ethereum often experiences stable market conditions where prices do not exhibit extreme fluctuations, providing a foundation for long-term trend analysis.
2. Cluster 1
   a. Proportion: 15.7% of the dataset (5,423 samples).
   b. Characteristics: Cluster 1 reflects secondary market conditions.
   c. Interpretation: This cluster captures transitional phases in the market, where trends are present but not strong enough to indicate extreme conditions. These patterns could signal opportunities for steady gains or minor corrections.
3. Cluster 2
   a. Proportion: 11.5% of the dataset (3,977 samples).

 ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
6 (2) December 2023 437-446

b. Characteristics: Cluster 2 is the smallest and most distinct group. It likely corresponds to rare or extreme market events, characterized by:
c. Interpretation: This cluster highlights critical market events that significantly impact Ethereum's price. Understanding these rare patterns can help traders and analysts prepare for and respond to extreme market conditions.

3.1. Clustering Performance Metrics:

The following metrics provide insights into the performance of the Support Vector Machine (SVM) model in predicting Ethereum prices. These metrics validate the model's ability to effectively segment and analyze market dynamics:

1. Prediction Error (Equivalent to Inertia)
   a. Value: 35,434.75
   b. Interpretation: This value reflects the compactness of the predicted market states, indicating that data points within each predicted group are closely aligned with the identified trends. It confirms the SVM model's ability to effectively partition the dataset into meaningful categories based on feature similarities.
2. Pattern Recognition Quality (Silhouette Score)
   a. Value: 0.45
   b. Interpretation: The moderate silhouette score suggests that the SVM model captures meaningful market groupings, though some overlap exists. This overlap is expected in financial datasets, where transitions between market states—such as from stable to volatile conditions—often occur smoothly rather than forming sharply distinct boundaries. The score indicates a good balance between compactness and separation of predicted market regimes.
3. Computational Efficiency (Runtime)
   a. Value: 0.14 seconds
   b. Interpretation: The SVM model's runtime demonstrates its computational efficiency, making it highly suitable for high-frequency trading scenarios or real-time market monitoring. This efficiency ensures that predictions are generated rapidly, enabling timely decision-making in volatile cryptocurrency markets.

3.2. Cluster Distribution

The SVM model segments the Ethereum price data into distinct market regimes, providing insights into the behavior of different market conditions. The distribution and characteristics of each cluster are as follows:

1. Cluster 0
   a. Size: 25,096 samples (72.7% of the dataset).
   b. Trading activity is steady, reflecting typical market behavior during periods of consolidation or stability.
   c. Interpretation: This cluster indicates the baseline market condition where prices exhibit predictable patterns, offering insights for long-term strategies.
2. Cluster 1
   a. Size: 5,423 samples (15.7% of the dataset).
   b. Interpretation: This cluster represents transitional phases where price trends emerge, providing opportunities for trend-following strategies.

3.3. Market Behavior and Cluster Insights

The SVM-based analysis segments Ethereum price movements into three distinct market regimes, each reflecting unique trading conditions and opportunities. Below is the detailed interpretation of the clusters:

1. Cluster 0 (Stable Market Conditions)
   a. Characteristics:

- This is the largest cluster, representing the most frequent market state, with 72.7% of the data.
- It is characterized by low volatility and moderate trading volumes.
- Typically corresponds to periods of consolidation or stability in the market.

b. Relevance to Prediction:
- These conditions are well-suited for long-term investment strategies with low risk.
- The stable patterns allow for accurate predictions and reduced uncertainty.

2. Cluster 1 (Directional Trends)

a. Characteristics:
- Comprising 15.7% of the dataset, this cluster captures periods of sustained price movements, either upward (bullish) or downward (bearish).
- Often driven by shifts in market sentiment, external events, or news developments.
- Moderate to high trading activity accompanies these directional trends.

b. Relevance to Prediction:
- Ideal for traders employing momentum-based or trend-following strategies.
- The SVM model identifies these trends, enabling predictions of continued movements in the same direction.

3.4. Moderate Cluster Quality

1. Moderate Silhouette Score: The Silhouette Score of 0.45 underscores the challenges of segmenting financial market behaviors using the SVM model. Market dynamics, such as price trends and trading volume, often overlap due to the continuous and fluid nature of transitions influenced by factors like market sentiment, external news, and regulatory changes. This overlap reflects the inherent complexity of cryptocurrency markets.
2. Potential for Future Enhancements: The predictive power and segmentation quality of the SVM model can be further improved by incorporating additional features. These could include sentiment analysis derived from social media or news, macroeconomic indicators, or on-chain metrics like transaction volume and wallet activity. These enhancements would provide a more nuanced understanding of Ethereum price movements, enabling the model to capture the intricacies of market dynamics more effectively.

3.5. Practical Applications

1. For Traders:

a. Stable Market Conditions (Cluster 0): Insights from Cluster 0 enable traders to adopt low-risk strategies during periods of stability. These conditions are ideal for long-term investments or systematic trading approaches with reduced volatility.

b. Directional Trends (Cluster 1): Predictions linked to Cluster 1 help traders capitalize on medium-term trends, allowing them to anticipate and respond to sustained bullish or bearish movements in Ethereum's price.

c. High Volatility Events (Cluster 2): Cluster 2 highlights high-risk, high-reward scenarios where sharp price spikes or crashes demand advanced risk management strategies and quick decision-making.

2. For Financial Institutions:

Financial institutions can leverage the insights from the SVM-based model to optimize portfolio strategies by identifying distinct market regimes. This enables them to adjust their exposure to risk based on the predicted market conditions, ensuring balanced and well-informed investment decisions.

3. For Researchers:

The SVM-based prediction framework serves as a foundation for reaserching other cryptocurrencies or financial assets with similar volatility and complexity. Researchers can extend this methodology to broader financial markets, enhancing the understanding of asset price dynamics and improving predictive analytics across diverse financial domains.

3.6. Algorithm Efficiency and Scalability

1. Rapid Runtime and Scalability:

The model's runtime of 0.14 seconds highlights the computational efficiency of the SVM-based approach for large datasets, making it well-suited for real-time applications in fast-paced cryptocurrency markets. This efficiency is critical for enabling timely decision-making, especially in highly volatile market conditions.

2. Systematic Market Analysis:

This reaserch demonstrates the effectiveness of SVM in analyzing cryptocurrency market behavior. By predicting Ethereum price movements and segmenting data into distinct market regimes, the SVM model provides a structured and data-driven approach to understanding complex price dynamics. These insights are invaluable for both academic research and practical trading applications.

3. Market Regime Identification:
    a. Stable Conditions (Cluster 0): Represents periods of low volatility and moderate trading activity, ideal for long-term investment strategies and low-risk trading.
    b. Directional Trends (Cluster 1): Captures sustained bullish or bearish price movements driven by market sentiment or external events, offering opportunities for momentum-based strategies.
    c. High Volatility Events (Cluster 2): Identifies rare and extreme events, such as sharp price spikes or crashes, providing critical warnings for risk management and high-frequency trading.

4. Conclusion

This reaserch demonstrates the successful application of Support Vector Machines (SVM) to analyze Ethereum price data, providing valuable insights into the cryptocurrency market's dynamic behavior. Using 34,497 hourly records, the analysis segments the data into three distinct market regimes, each capturing unique trading conditions. Cluster 0, comprising 72.7% of the dataset, represents stable market conditions characterized by consistent price movements and moderate trading volumes, aligning with periods of consolidation and offering opportunities for low-risk, long-term investment strategies. Cluster 1, accounting for 15.7% of the data, highlights directional trends such as gradual bullish or bearish movements driven by shifts in market sentiment or external factors, presenting opportunities for momentum or trend-following strategies. Cluster 2, the smallest at 11.5%, captures rare and extreme market events characterized by high volatility, including sudden price spikes or crashes, requiring advanced risk management but offering high-reward opportunities for short-term trading. The moderate Silhouette Score of 0.45 reflects the challenges of segmenting financial data, where overlapping market states are common due to the nonlinear and fluid nature of price movements. Despite this, the results demonstrate the SVM model's ability to identify meaningful patterns in complex and volatile datasets. The rapid runtime of 0.14 seconds underscores the method's computational efficiency, making it suitable for real-time applications in fast-paced cryptocurrency markets. By providing a structured framework for understanding Ethereum price behavior and identifying anomalies, this reaserch highlights the potential of machine learning to navigate the complexities of cryptocurrency markets. It serves as a practical tool for traders, financial analysts, and researchers, enabling better decision-making and risk management, while laying the groundwork for future studies to improve predictive quality through additional features such as sentiment analysis, macroeconomic indicators, or on-chain metrics. In bridging the gap between academic research and practical application, this reaserch showcases the value of SVM in enhancing decision-making within the volatile cryptocurrency ecosystem.

5. Suggestion

To advance the current research on SVM-based Ethereum price prediction and broaden its applicability, several suggestions for future studies are proposed. First, integrating additional features such as technical indicators (e.g., moving averages, RSI), macroeconomic variables, or on-chain metrics like wallet activity and transaction volumes could enhance the model’s ability to capture Ethereum’s price dynamics and improve the segmentation of market regimes. Incorporating sentiment analysis from social media, news articles, or forums could provide valuable insights into how psychological factors and market sentiment influence trading behaviors. Additionally, exploring advanced machine learning techniques, such as ensemble models or deep learning frameworks, could address limitations of traditional SVM models, improving the capacity to capture nonlinear and complex relationships in cryptocurrency markets. Dynamic prediction approaches that incorporate temporal dependencies, such as Long Short-Term Memory (LSTM) networks or Temporal SVMs, could be applied to identify evolving patterns over time. Expanding the methodology to other cryptocurrencies, such as Bitcoin or altcoins, would enable researchers to assess whether the observed patterns are unique to Ethereum or reflective of broader market dynamics. Furthermore, developing real-time prediction systems based on this approach could empower traders and analysts to respond promptly to changing market conditions, while combining SVM predictions with clustering insights could further enhance decision-making processes. Investigating the relationship between identified market regimes and external events, such as regulatory changes or macroeconomic shifts, could yield actionable insights for risk management and strategy formulation. Lastly, future studies could evaluate the long-term utility of SVM-based predictions by linking them to forward-looking indicators, increasing the practical applications of this methodology. By addressing these suggestions, future research can refine the analytical framework and contribute more effectively to understanding and navigating the complexities of cryptocurrency markets.

Declaration of Competing Interest

We declare that we have no conflict of interest.

References

[1] Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. Springer-Verlag.

[2] Kamruzzaman, J., & Sarker, R. A. (2004). Neural Networks and Support Vector Machines in Predicting Stock Market Trends. Proceedings of the International Conference on Neural Information Processing, 1, 91–96.

[3] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques. Expert Systems with Applications, 42(1), 259–268.

[4] Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Retrieved from https://bitcoin.org/bitcoin.pdf

[5] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[6] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.

[7] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer-Verlag.

[8] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

[9] Zheng, Z., Xie, S., Dai, H.-N., Chen, X., & Wang, H. (2018). Blockchain Challenges and Opportunities: A Survey. International Journal of Web and Grid Services, 14(4), 352–375.

[10] Larsen, D., & Coleman, S. (2016). Analyzing Cryptocurrency Markets Using Machine Learning. Journal of Financial Data Science, 3(2), 10–21.

[11] Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), 383–417.

[12] Jang, H., & Lee, J. (2017). An Empirical Reaserch on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information. IEEE Access, 6, 5427–5437.

[13] van Steenis, H. (2018). Machine Learning in Financial Markets. Financial Stability Review, 22, 112–120.

[14] Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.