Faktor Exacta

Vol. 18, No. 1, March 2025, pp. 96~103
eISSN: 2502-339X, pISSN: 1979-276X, DOI: https://doi.org/10.30998/faktorexacta.v18i1.27805 ❒ 96

Identification of Stock Breakouts Using Support Vector Machine with Integrated Fundamental Data and Random Forest Prediction

Gusti Bagus Cahya Utama[1], Ahmad Chusyairi[2], Riad Sahara[3]
[1,2,3]PJJ Informatika, Universitas Siber Asia, Indonesia

Article Info

Article history:

Received Sep 9, 2024
Revised May 20, 2025
Accepted Jun 11, 2025

Keywords:

Breakout Identification
Fundamental data
Random Forest (RF)
Stock market analysis
Support Vector Machine (SVM)

ABSTRACT

In this study, we investigated the detection of breakout events in Tesla, Inc. stock by integrating technical analysis with fundamental financial data using a Support Vector Machine (SVM) model. The Opening Range Breakout (ORB) strategy has demonstrated substantial returns, while the Support Vector Machine (SVM) method excels in detecting breakout events. Furthermore, the Random Forest (RF) algorithm effectively forecasts long-term trends. This study aims to integrate fundamental data—specifically net income and Earnings Per Share (EPS)—and a long-term trend prediction derived from RF as additional features in an SVM model for Tesla, Inc. (TSLA) stock. Utilizing the Sample, Explore, Modify, Model, Assess (SEMMA) framework, the research evaluates daily stock data from November 15, 2019, to November 14, 2024. Results indicate that incorporating fundamental data improves SVM precision from 0.08 to 0.18, although recall remains low. Conversely, adding the RF prediction feature does not yield a significant benefit and reduces precision to 0.13. These findings suggest that while integrating fundamental data enhances breakout detection performance, further refinement is essential for the effective incorporation of RF-based predictions.

Corresponding Author:

Ahmad Chusyairi,
PJJ Informatika,
Universitas Siber Asia,
Jl. RM. Harsono, Ragunan - Jakarta Selatan.Daerah Khusus Ibukota Jakarta 12550
Email: ahmadchusyairi@lecturer.unsia.ac.id

1. INTRODUCTION

The Stock trading in financial markets offers significant potential for financial gains; however, the inherent complexity of stock behaviour demands comprehensive analysis [1]. Prior research has demonstrated that the Opening Range Breakout (ORB) strategy can yield substantial returns, with net gains exceeding 1600% when applied to highly traded stocks [2]. In this context, predictive models based on machine learning have emerged as valuable tools for identifying breakout events in stock markets.

Recent studies have highlighted the superior performance of Support Vector Machine (SVM) models in detecting stock breakouts compared to other machine learning approaches [3], [4], [5]. Concurrently, the Random Forest (RF) algorithm has proven effective in forecasting long-term trends [6]. Despite these advancements, previous investigations have predominantly relied on technical indicators, thereby overlooking the potential benefits of incorporating fundamental data—such as net income and Earnings Per Share (EPS)—which provide a broader contextual understanding of price movements [7].

The United States stock market, characterized by its sustained efficiency even during periods of high volatility, presents an ideal environment for data-driven approaches. High daily trading volumes—such as the

approximately 8.26 billion shares traded on NASDAQ on November 11, 2024, with a transaction value of around $359.27 billion—further underscore the suitability of this market for rigorous empirical analysis [8], [9]. Tesla, Inc. (TSLA) is selected as the focus of this study due to its impressive financial performance and pronounced price volatility [10], [11].

This research focuses on the identification of breakout events in TSLA stock using an SVM framework. The proposed model integrates not only technical indicators—including Average True Range (ATR), Bollinger Bands Width, Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD)—but also fundamental financial data and long-term trend predictions derived from the RF algorithm. The integration of fundamental data and RF predictions is expected to enhance the accuracy of breakout detection, thereby providing a more robust tool for investors and contributing to the academic literature on stock prediction.

In summary, this study aims to address the following research questions:

1. What is the impact of including fundamental data on the accuracy of SVM models for detecting stock breakouts?
2. How effective is the integration of long-term trend predictions from the RF algorithm as an additional feature for SVM models in breakout identification?

By leveraging a systematic approach based on the Sample, Explore, Modify, Model, Assess (SEMMA) methodology, this research seeks to develop a comprehensive predictive framework that effectively captures the multifaceted dynamics of stock price movements.

2. METHOD

This study applies the Sample, Explore, Modify, Model, Assess (SEMMA) framework to systematically guide data analysis and model construction, given its recognized flexibility in handling both technical and fundamental stock variables [12], [13]. By structuring the research under SEMMA, the methodology ensures that data sampling is performed efficiently, exploration uncovers critical insights, modification refines features, modeling applies a chosen predictive algorithm, and assessment gauges the resulting performance against relevant metrics [14].

Figure 1 presents the application of the SEMMA methodology (Sample, Explore, Modify, Model, Assess) to the analysis of Tesla stock data. In the Sample phase, daily stock data of Tesla shares traded on the U.S. stock market are selected. The Explore phase involves a comprehensive analysis of technical and fundamental data to discern patterns and trends associated with stock breakout behavior. During the Modify phase, additional features, such as long-term trend predictions using Random Forest and the integration of fundamental data, are incorporated to enhance the dataset. In the Model phase, a Support Vector Machine (SVM) is employed to detect breakout events by combining technical and fundamental characteristics. Finally, the Assess phase focuses on evaluating the model's performance, with particular emphasis on the accuracy of breakout identification throughout the study period.

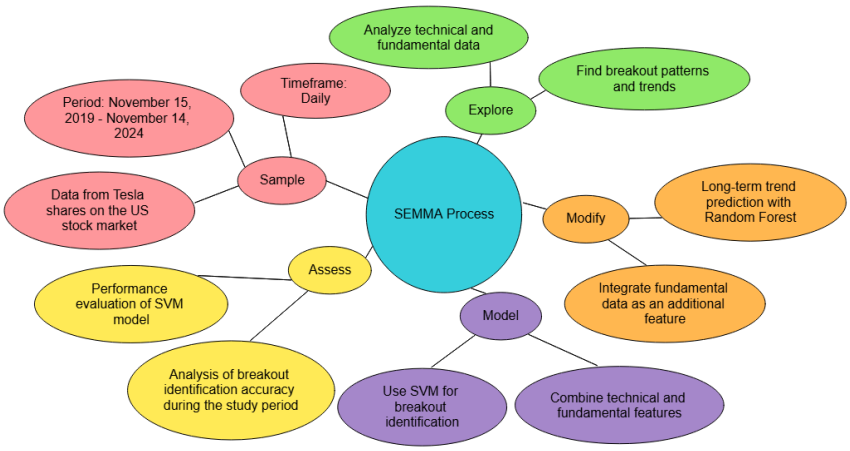

Figure 1. SEMMA

2.1. Sample

Data were collected from daily Tesla (TSLA) stock records between November 15, 2019, and November 14, 2024, emphasizing representativeness in the sampling process. The chosen timeframe coincides with the maximum retrieval capacity of the Alpha Vantage API for fundamental data, ensuring a comprehensive view of Tesla's performance. Each record comprises date, open price, high price, low price, close price, and trading volume. Missing or anomalous values were addressed using statistical detection methods such as z-score and interquartile range (IQR) analysis, followed by linear interpolation to preserve the continuity of data [15].

2.2. Explore

Data exploration centered on identifying patterns within technical indicators and fundamental variables that could signal potential breakout events. Technical indicators included Average True Range (ATR), Bollinger Bands Width (BB_Width), Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD). Fundamental attributes encompassed net income and Earnings Per Share (EPS), supplemented by the calculation of percentage changes (pct_netIncome and pct_eps) [16], [17]. A breakout was labeled if the closing price on a given day exceeded the highest price recorded during the preceding ten trading days [3]. This phase also entailed detecting outliers and investigating correlations among variables to improve feature selection.

2.3. Modify

Feature engineering was conducted by adding a 90-day trend forecast from a Random Forest (RF) model trained on open, high, low, close, and volume data. This forecast was encoded as a binary variable (rf_pred_up_90d) to signify expected upward price movement or otherwise [6]. All numerical inputs were then normalized via StandardScaler to ensure that the SVM model treated each feature on an equivalent scale, thus enhancing separation in the high-dimensional feature space [13]. The integration of fundamental metrics enriched the technical analysis, thereby facilitating a more comprehensive understanding of Tesla's corporate performance and its consequent impact on share price movements.

2.4. Model

The core predictive engine was constructed using Support Vector Machine (SVM), selected for its efficacy in detecting breakout patterns [3]. Configured with the Radial Basis Function (RBF) kernel, the SVM model leverages parameters such as gamma (0.1) to regulate the influence radius of single data points and nu (0.05) to set the proportion of outliers permissible in the training process. By combining technical indicators, fundamental variables, and the RF-generated trend feature, the resulting SVM sought to capture both short-term market fluctuations and longer-term performance signals.

2.5. Assess

Evaluation of the predictive model was conducted using performance metrics such as accuracy, precision, recall, and F1-score, which have been widely recognized for assessing the reliability of stock breakout predictions [18]. Accuracy, defined as the proportion of correct predictions, served as a general indicator of model performance and was particularly important when evaluating datasets with relatively balanced class distributions [19]. Precision measured the correctness of the positive breakout signals, while recall gauged the model's ability to detect all actual breakout events; the F1-score, as the harmonic mean of precision and recall, provided a balanced assessment especially in contexts with imbalanced data [20], [21]. Comparative testing was performed across three different approaches: a technical-only model, a hybrid model that integrated both technical indicators and fundamental data, and a fully integrated model that also incorporated a 90-day trend prediction derived from a Random Forest algorithm. In line with prior research, the consistency and stability of the model behavior were examined to determine whether the addition of fundamental data and the RF-derived trend feature enhanced the predictive capability or instead introduced additional noise, thereby affecting the detection of breakout events [6].

Table 1. The Performance of SVM Breakouts Detection

Model	Accuracy	Precision	Recall	F1-Score
Technical data only	0.89	0.08	0.03	0.04
Technical and fundamental data	0.89	0.18	0.05	0.08
Technical data, fundamentals, and RF prediction	0.89	0.13	0.04	0.06

This table indicates that although the overall accuracy remains unchanged, the incorporation of supplementary data—such as fundamental data and RF predictions—can lead to improvements in other

evaluation metrics, namely precision, recall, and F1-score, which are pertinent in the context of breakout prediction. Nevertheless, the SVM model that integrates RF predictions shows a decline in performance regarding precision, recall, and F1-score compared to the model augmented solely with fundamental data.

3. RESULT AND DISCUSSION

The predictive models developed in this study were evaluated on a daily Tesla (TSLA) stock dataset, focusing on three primary approaches: using only technical data, combining technical and fundamental data, and further incorporating the Random Forest (RF) long-term trend feature. All three approaches yielded an accuracy of 0.89, indicating that the models performed consistently in correctly classifying a substantial portion of the data overall. However, the key metrics for assessing breakout detection—namely precision, recall, and F1-score—displayed notable variation across the three approaches.

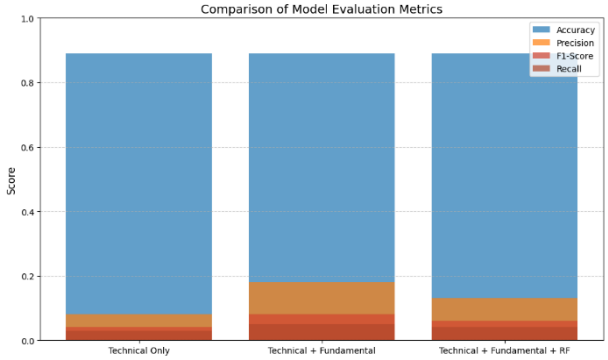

Figure 2. Comparison of Evaluation Metrics Across Predictive Model Variants

3.1. Performance of Technical Indicators and Impact of Fundamental Data in Breakout Detection

In the first approach, which employed only technical indicators such as Average True Range, Bollinger Bands Width, Relative Strength Index, and Moving Average Convergence Divergence, the model exhibited a precision of 0.08 and a recall of 0.03. These results reveal that although the accuracy was high, the model's ability to correctly identify actual breakouts remained limited, as evidenced by a low F1-score of 0.04.

Figure 2 further depicts this limitation by showing that, between September 2, 2020, and several months thereafter, the SVM model based exclusively on technical data failed to detect any breakouts in the observed period, even though there were several significant upward price movements. This outcome underscores the model's inadequacy in recognizing genuine breakout patterns during that time frame. The deficiency is primarily attributed to the model's reliance on technical indicators that are insufficiently sensitive to certain price movements, thereby impeding its ability to capture critical signals of a breakout.

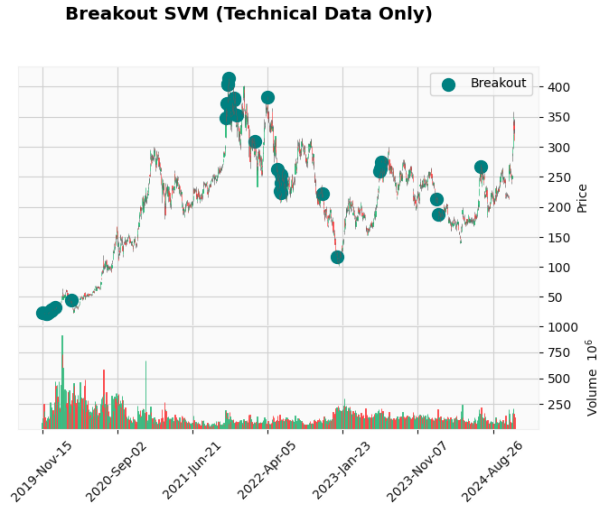

Figure 3. SVM Breakout Identification Chart only with technical data

The second approach, which integrated fundamental information (net income and Earnings Per Share) into the technical indicators, improved the precision from 0.08 to 0.18, representing a relative increase

of 125%, indicating a significant improvement in the model's ability to classify breakout events. However, the recall value remained at 0.05, indicating that while fewer false positives emerged, the model still struggled to capture the total number of true breakouts.

The chart in Figure 3 demonstrates that with the integration of both technical and fundamental data, the SVM model is able to detect more breakouts compared to the model that uses only technical data. This improvement shows that fundamental data, such as net income and EPS, provides additional information that strengthens the model's capability to identify significant stock price movement patterns. However, there are still some misses in breakout detection, especially in price areas reaching all-time highs. This occurs because the model tends to assume that a new highest price always constitutes a breakout, even though in some cases the price movement is not followed by a pattern that truly reflects a breakout. This issue indicates that, although the integration of fundamental data is beneficial, the model still needs to be refined to recognize more complex patterns and reduce erroneous assumptions.

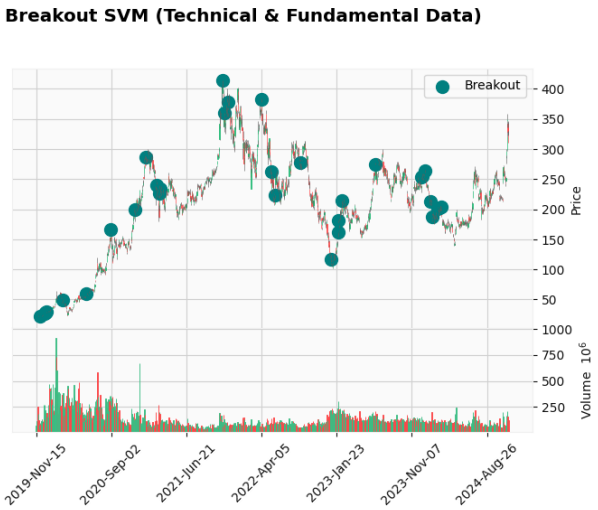

Figure 4. SVM Breakout Identification Chart with the Integration of Fundamental Data

3.2. The Impact of Random Forest Long-Term Trend Predictions

When the Random Forest long-term trend feature was added to the technical and fundamental inputs, the model's precision declined from 0.18 to 0.13, representing a relative decrease of 27.8%, and the recall dropped from 0.05 to 0.04, representing a relative decrease of 20%. This decline suggests that the simplistic binary encoding of the RF trend prediction introduced additional noise, reducing the model's capacity to consistently detect true breakouts. Although the model became more conservative in labeling potential breakout points, this greater caution did not translate into an overall improvement in predictive performance, as evidenced by a lower F1-score of 0.06 compared to the 0.08 achieved by the technical-plus-fundamental model.

SVM model becomes more selective on detecting breakouts, as evidenced in Figure 4. Some breakout points that were detected by previous models, such as in the January 2023 area, are no longer identified by the model. This indicates that the model has become more cautious in declaring a breakout, due to the additional information from the 90-day trend prediction provided by the Random Forest. However, this approach also results in a decrease in precision, as the model begins to lose the ability to consistently detect valid breakouts. Although the model becomes more vigilant against false breakout signals, this increased selectivity actually introduces noise that diminishes precision in breakout detection.

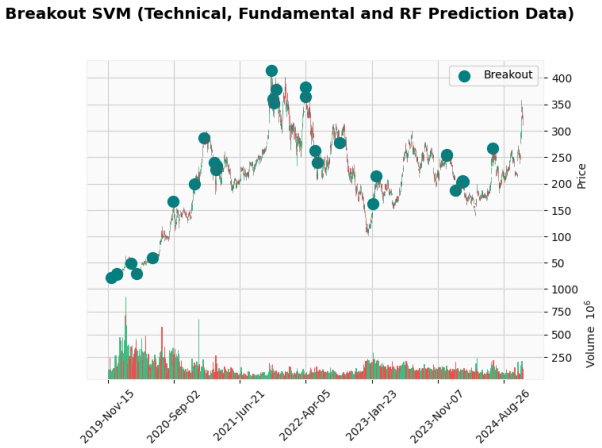

Figure 5. SVM Breakout Detection Chart with the Integration of the Random Forest Trend Prediction Feature

3.3. Limitations of Accuracy in Minority Class Evaluation

Despite the observed discrepancies in precision and recall, the uniform accuracy of 0.89 across all approaches highlights the importance of looking beyond a single evaluation metric, particularly when the event of interest (i.e., breakout) occurs relatively infrequently. A high accuracy can mask poor performance on such minority classes, underscoring the need for metrics that more directly capture the model's discrimination capacity. In this context, the integration of fundamental data proved beneficial by adding contextual richness to price movements, although the incorporation of long-term trend predictions did not yield an equivalent advantage. These findings align with existing literature suggesting that while machine learning techniques can enhance breakout detection, data representation and feature engineering remain critical to achieving robust performance.

Overall, the results illustrate both the promise and the limitations of using combined technical and fundamental data within a Support Vector Machine framework. Fundamental data integration improved precision, whereas the Random Forest-based feature did not deliver a tangible increase in predictive quality, likely due to the binary nature of its implementation. Future research could explore alternative representations of the RF output, such as confidence intervals or probabilistic values, to reduce noise and capture more nuanced trend information. By refining breakout definitions and incorporating broader data sources—potentially including macroeconomic indicators, market sentiment, and multi-timeframe analyses—the model may achieve higher recall rates and better generalization to other stocks or market conditions.

4. CONCLUSION

This research investigated how integrating fundamental data and Random Forest (RF) long-term predictions within a Support Vector Machine (SVM) framework could enhance the detection of breakout events in Tesla (TSLA) stock. Although all tested models—technical only, technical plus fundamental, and technical plus fundamental plus RF—achieved a uniform accuracy of 0.89, notable differences emerged in their ability to correctly identify breakout instances. Incorporating fundamental variables (net income and Earnings Per Share) substantially improved precision from 0.08 to 0.18, underlining the value of augmenting technical analysis with corporate performance data. However, the binary RF-based feature did not produce a corresponding benefit, as its inclusion introduced noise that reduced precision to 0.13. These findings demonstrate the need to go beyond simple accuracy metrics when assessing stock breakout models, given the relative rarity of breakout events. Additionally, they highlight the potential of fundamental data to provide meaningful context for technical signals, as well as the importance of careful feature engineering for long-term trend forecasts. Future work can refine the RF prediction feature by using probabilistic outputs instead of binary labels, extend the dataset to include macroeconomic indicators or sentiment analysis, and perform more robust parameter tuning to address the persistent challenge of low recall.

Integration of fundamental data such as net income and Earnings Per Share (EPS) is proven to increase the precision of the SVM model in detecting TSLA stock breakouts. The technical data-based model alone has a precision of 0.08, while with the addition of fundamental data, the precision increases to 0.18. This suggests that the fundamental data provides a richer context to the stock price movement, so the model can reduce false breakout predictions. However, recall remains low, the overall accuracy of the model has not shown a significant improvement.

The addition of long-term trend prediction (90 days) from the Random Forest algorithm does not provide a significant performance improvement to the SVM model on TSLA stocks. Precision of the model actually decreased from 0.18 to 0.13 when Random Forest prediction results were added in the SVM model for breakout identification. This indicates that the additional trend prediction feature from Random Forest introduces noise or less relevant information. Therefore, there is a need to re-evaluate the way the Random Forest prediction features are represented in the SVM model. One approach that could be considered is to use the probability or confidence index of Random Forest predictions as a feature for the SVM model in breakout identification, instead of binary values (1 and 0) to ensure the information added is more meaningful and relevant to breakout identification.

REFERENCES

[1] R. Rohyati, F. P. N. Rokhmah, H. N. U. Syazeedah, R. I. Fitriyaningrum, G. Ramadhan, and M. Syahwildan, "Tantangan dan Peluang Pasar Modal Indonesia dalam Meningkatkan Minat Investasi di Era Digital," *Kompeten: Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 3, no. 1, pp. 909–918, 2024, doi: 10.57141/kompeten.v3i1.133.

[2] C. Zarattini, A. Barbon, and A. Aziz, "A Profitable Day Trading Strategy For The U.S. Equity Market," *SSRN Electronic Journal*, pp. 24–98, 2024, doi: 10.2139/ssrn.4729284.

[3] Md. S. Ansary, "Breakout Stocks Identification using Machine Learning Approaches," *ENP Engineering Science Journal*, vol. 2, no. 2, pp. 52–56, 2022, doi: 10.53907/enpesj.v2i2.173.

[4] N. F. Silva, L. P. de Andrade, W. S. da Silva, M. K. de Melo, and A. O. Tonelli, "Portfolio optimization based on the pre-selection of stocks by the Support Vector Machine model," *Financ Res Lett*, vol. 61, p. 105014, 2024, doi: 10.1016/j.frl.2024.105014.

[5] X. Chen and Z.-J. He, "Prediction of Stock Trading Signal Based on Support Vector Machine," in *2015 8th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, IEEE, 2015, pp. 651–654. doi: 10.1109/ICICTA.2015.165.

[6] J. Zheng, D. Xin, Q. Cheng, M. Tian, and L. Yang, "The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance," *Atlantis Highlights in Computer Sciences*, vol. 21, pp. 82–90, 2024, doi: 10.2991/978-94-6463-419-8_11.

[7] Y. Chen and T.-T. Hsu, "Why US Stock Markets Have Recovered So Fast from the Pandemic Crash," *Universal Journal of Accounting and Finance*, vol. 9, no. 5, pp. 972–981, 2021, doi: 10.13189/ujaf.2021.090508.

[8] H. Hong, Z. Bian, and C.-C. Lee, "COVID-19 and instability of stock market performance: evidence from the U.S.," *Financial Innovation*, vol. 7, no. 1, p. 12, 2021, doi: 10.1186/s40854-021-00229-1.

[9] NasdaqTrader, "Daily Market Summary." Accessed: Nov. 16, 2024. [Online]. Available: https://www.nasdaqtrader.com/Trader.aspx?id=DailyMarketSummary

[10] Z. Wei, "A Financial Analysis and Valuation of Tesla," *Advances in Economics, Management and Political Sciences*, vol. 102, no. 1, pp. 254–259, 2024, doi: 10.54254/2754-1169/102/2024ED0086.

[11] O. Dovbnya, "Riding the Bull and the Bear: A Metaphor Analysis of Tesla's Stock Volatility in Business Media Discourse," *Scientific Journal of Polonia University*, vol. 58, no. 3, pp. 50–57, 2023, doi: 10.23856/5807.

[12] W. Yustanti, N. Iriawan, and I. Irhamah, "Categorical encoder based performance comparison in pre-processing imbalanced multiclass classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, p. 1705, 2023, doi: 10.11591/ijeecs.v31.i3.pp1705-1715.

[13] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: a systematic literature review," *PeerJ Comput Sci*, vol. 6, p. e267, 2020, doi: 10.7717/peerj-cs.267.

[14] Omari Firas, "A combination of SEMMA & CRISP-DM models for effectively handling big data using formal concept analysis based knowledge discovery: A data mining approach," *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 009–014, 2023, doi: 10.30574/wjaets.2023.8.1.0147.

[15] B. Bicski and A. Pekar, "Unveiling Latency-Induced Service Degradation: A Methodological Approach With Dataset," *IEEE Access*, vol. 12, pp. 128097–128116, 2024, doi: 10.1109/ACCESS.2024.3456588.

[16] R. N. Aisyah, R. Damayanti, E. Lilianti, P. Manajemen, F. Ekonomi, and D. Bisnis, "Pengaruh Laba Bersih dan Arus Kas terhadap Harga Saham pada Perusahaan Sub Sektor Kimia yang Terdaftar di Bursa Efek Indonesia," *Journal of Management Small and Medium Enterprises (SME's)*, vol. 16, no. 3, pp. 531–539, 2023.

[17] D. N. H. Ilahiyah, "Pengaruh Earning Per Share (EPS) dan Pertumbuhan Penjualan terhadap Harga Saham pada Perusahaan Farmasi yang Terdaftar di Bursa Efek Indonesia (BEI)," *Akuntansi Dewantara*, vol. 5, no. 2, 2021, doi: 10.26460/ad.v5i2.9574.

[18] T. Hidayati, D. Wulandari, and W. G. Aedi, "Implementasi Algoritma C4.5 Dalam Memprediksi Harga Saham," *Scientia Sacra: Jurnal Sains*, vol. 3, no. 4, pp. 1–7, 2023.

[19] M. Harju and A. Mesaros, "Evaluating Classification Systems Against Soft Labels with Fuzzy Precision and Recall," *Detection and Classification of Acoustic Scenes and Events 2023*, 2023, doi: https://doi.org/10.48550/arXiv.2309.13938.

[20] M. Khayatkhoei and W. AbdAlmageed, "Emergent Asymmetry of Precision and Recall for Measuring Fidelity and Diversity of Generative Models in High Dimensions," *Proceedings of the 40th International Conference on Machine Learning*, 2023, doi: https://doi.org/10.48550/arXiv.2306.09618.

[21] S. Riyanto, I. S. Sitanggang, T. Djatna, and T. D. Atikah, "Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.01406116.