Research Horizon

ISSN: 2808-0696 (p), 2807-9531 (e)

Research Horizon
Volume: 06
Issue: 03
Year: 2026
Page: 1375-1386

Citation:
Sitohang, L., Sitohang, S., & Soetanto, H. (2026). A comparative analysis of machine learning regression models for TikTok user engagement prediction. *Research Horizon*, *6*(3), 1375-1386.

Article History:
Received: May 17, 2026
Revised: June 11, 2026
Accepted: June 24, 2026
Online since: June 25, 2026

A Comparative Analysis of Machine Learning Regression Models for TikTok User Engagement Prediction

Lidiasonata Sitohang[1*], Swinsikya Sitohang[1], Hari Soetanto[1]

[1] Department of Computer Science, Faculty of Information Technology, Universitas Budi Luhur, Jakarta, Indonesia

[*] *Corresponding author: Lidiasonata Sitohang (sonatalidiasit@gmail.com)*

Abstract

The rapid growth of TikTok has made user engagement prediction a critical challenge for content creators and digital marketers, particularly given the high multicollinearity among interaction features such as likes, comments, and shares. This study aims to conduct a comparative analysis of three machine learning models, namely linear regression, elastic net, and support vector regression, in predicting TikTok user engagement levels. The methodology employs a quantitative approach using the cross-industry standard process for data mining framework, evaluating model performance through mean absolute error, root mean squared error, mean absolute percentage error, and coefficient of determination metrics. Findings reveal that the elastic net is the most reliable model, achieving a mean absolute error of 3.98 and root mean squared error of 9.37 with a coefficient of determination of 1.000, supported by consistent cross-validation results across five folds. Linear regression produced trivial perfect scores due to the direct summation relationship between input features and the target variable, while support vector regression demonstrated suboptimal performance with a mean absolute error of 74.58, indicating difficulty in capturing linear data patterns. These results suggest that regularization-based models offer a more practical and generalizable framework for social media engagement prediction, providing actionable insights for practitioners in developing data-driven content strategies.

Keywords

Elastic Net, Engagement Prediction, Linear Regression, Machine Learning, Predictive Modeling, SVR, TikTok.

1. Introduction

In today's era of digital transformation, social media platforms have undergone significant evolution, with TikTok emerging as the global market leader in short-form video. The platform's rapid growth has reshaped how content creators, brands, and digital marketers communicate with their audiences, making engagement metrics the most critical indicator of content success (Erlany et al., 2022). The platform's massive growth is driven by its highly personalized recommendation algorithm, making user engagement metrics the most crucial indicator of success for creators and digital marketers. However, the volatility of trends and the sheer volume of data create challenges for practitioners in accurately predicting content performance. This uncertainty often leads to inefficient marketing strategies, suboptimal resource allocation, and missed opportunities for the audience (Xiao et al., 2026).

Predicting user engagement on TikTok has attracted growing attention from researchers in recent years. Several studies have attempted to model engagement behavior using various machine learning approaches. For instance, Wanajma (2024) applied support vector regression and random forest to predict engagement rate on TikTok, while Safrin and Simanjorang (2023) compared multiple machine learning algorithms for the same purpose. Similarly, Talebi and Abdolvand (2025) utilized ensemble learning with likes, comments, and shares as predictor variables to estimate TikTok engagement rate. These studies collectively highlight that interaction-based features such as likes, comments, and shares are strong predictors of engagement performance (Putri & Hendrawan, 2026).

Beyond TikTok, engagement prediction has also been explored across other digital platforms. Gunawan and Suhendra (2022) compared random forest and support vector regression for predicting social media engagement rate in a broader context, while Asmawi et al. (2025) analyzed likes and comments features for social media interaction prediction using a random forest approach. Sapina et al. (2025) further demonstrated that algorithm selection significantly affects prediction accuracy when comparing random forest and SVR across social media datasets. These findings suggest that no single algorithm consistently outperforms others across all contexts, and that dataset characteristics play a major role in determining model effectiveness.

Despite extensive research on social media, there is a research gap in the effectiveness of regression algorithms for data characterized by multicollinearity and strong linear correlation, such as the TikTok dataset. According to John (2010), standard regression models often struggle when input features are highly correlated, as multicollinearity inflates coefficient variance and reduces model reliability. Standard regression models often fail to address the complexity of social media data, while more complex models sometimes suffer from overfitting. Wanajma (2024) also noted that virality factors on TikTok tend to exhibit high inter-variable correlation, which further complicates the modeling process. Safrin and Simanjorang (2023) similarly found that interaction features such as likes, comments, and shares are often linearly dependent, making it difficult for standard regression to isolate the individual contribution of each feature. There is an urgent need to compare regression models capable of automatically performing feature regularization to produce more stable and reliable predictions within the context of TikTok's audience dynamics.

This study aims to conduct a comparative analysis between three machine learning algorithms, namely linear regression, elastic net, and Support Vector Regression (SVR), in predicting TikTok content viewership. This study seeks to answer whether elastic net regularization can outperform standard linear regression and SVR in handling multicollinearity within TikTok interaction data. Through a quantitative approach, this study evaluates the performance of each model using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2 metrics. This research is expected to fill the existing gap by providing empirical evidence on the most suitable regression framework for TikTok engagement prediction, while also offering a practical reference for content analysts and digital marketing practitioners in making data-driven decisions.

2. Literature Review

2.1. Linear Regression in Predictive Modeling

Linear Regression is one of the most widely used algorithms in predictive modeling due to its simplicity and interpretability. Kuhn and Johnson (2013) and Ababil et al. (2022) applied linear regression to predict liquid vape sales and demonstrated that the method produces reliable estimates when the relationship between variables is predominantly linear. Similarly, Adiguno et al. (2022) utilized multiple linear regression to forecast sales turnover and found that the model performs consistently when input features maintain a stable linear trend. These studies collectively suggest that linear regression remains a practical baseline model for prediction tasks across various domains.

Several other studies have further confirmed the applicability of linear regression in diverse prediction contexts. Kwok and Susanti (2019) implemented linear regression for raw material demand forecasting in a tofu production system, while Padilah and Adam (2019) applied multiple linear regression to estimate rice productivity in Karawang Regency. Harsiti and Srihartini (2022) also demonstrated its effectiveness in predicting tablet medicine inventory, achieving stable prediction results with relatively low error rates. In the context of short-form video platforms, Smith and Mason (1997) and McCarthy et al. (2022) implemented linear regression within a point-of-sales restaurant application and found that direct linear relationships between variables tend to produce near-perfect model performance. Akhmad (2020) and Almumtazah et al. (2021) further reinforced this pattern, showing that linear regression consistently achieves high accuracy when target variables are directly derivable from input features, though this condition also raises concerns about the model capturing trivial rather than meaningful relationships.

2.2. Engagement and Content Popularity Prediction

Predicting user engagement and content popularity on short-form video platforms has become an active research area as platforms like TikTok continue to grow. Xiao et al. (2026) applied multiple linear regression to predict engagement rate for short videos and found that interaction-based features such as likes and comments are strong predictors of overall engagement performance. Li et al. (2024) further confirmed this by using multiple linear regression to estimate digital content popularity based on user interaction data, reporting consistent prediction accuracy across different content categories. These findings highlight the potential of regression-based approaches for modeling content performance on video platforms.

Fatimah and Nasir (2025) investigated the influence of likes and shares on short video popularity estimation and reported that these two features alone contribute significantly to prediction accuracy. May and Siddo (2024) compared linear regression and decision trees for content popularity estimation and found that linear regression performs competitively when the data structure is predominantly linear. Li et al. (2024) implemented both SVR and a random forest regressor for digital interaction prediction and highlighted that SVR tends to underperform when data follows a strictly linear distribution. Kim et al. (2024) analyzed the performance of multiple regression algorithms for predicting short video content popularity and emphasized that algorithm selection should be guided by the underlying data characteristics. Hardiyanto and Rozi (2020) and Almutairi and Rawat (2024) also noted that feature-based interaction data from social media platforms often contains redundant variables, which can negatively affect model generalization if regularization is not applied.

2.3. Engagement Prediction on YouTube and E-Commerce Platforms

Engagement and viewership prediction have also been extensively studied on YouTube and e-commerce platforms, providing useful comparative insights for TikTok-related research. Andariesta and Wasesa (2023) compared linear regression, random forest, and SVR for predicting YouTube viewer counts and found that algorithm performance varies significantly depending on the degree of linearity present in the dataset. Guo et al. (2024) conducted a comparative analysis of regression algorithms for YouTube video popularity prediction based on metadata and reported that simpler models sometimes outperform complex ones when feature-target relationships are straightforward. These results suggest that dataset characteristics, rather than model complexity alone, are the primary determinant of predictive performance.

Asante et al. (2023) estimated YouTube viewer counts using machine learning and found that interaction-based features remain consistently predictive across different content creator profiles in Indonesia. Irshad et al. (2024) compared regression algorithms for YouTube view count prediction and emphasized the importance of evaluating multiple metrics beyond R^2 to avoid misleading performance assessments. Johnson and Malaga (2024) further reinforced the value of ensemble and regression-based approaches for popularity prediction tasks. In the e-commerce domain, Wanajma (2024) demonstrated that machine learning models can reliably forecast customer engagement based on digital interaction data. Safrin and Simanjorang (2023) compared multiple regression models for predicting consumer engagement on e-commerce applications and concluded that no single model universally dominates, reinforcing the need for systematic comparative studies such as the one conducted in this research.

3. Methods

This research applies the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework as the overarching methodology, which includes business understanding, data understanding, data preparation, modeling, and evaluation (Murti, 2024). This framework was selected because it provides a structured and systematic pipeline that ensures each stage of the research process is well-defined and reproducible. The overall workflow of this framework is illustrated in Figure 1.

Figure 1. CRISP-DM Framework for Research Workflow

The dataset used consists of TikTok user interaction attributes, including the number of likes, comments, and shares, as input features, with total engagement as the target variable. The dataset contains records collected from publicly available TikTok content data, covering a range of content categories and creator profiles. Data preparation involves several crucial steps. Feature selection is conducted to determine independent variables that significantly influence the target engagement variable. Structure checking is performed to verify data types and clean the data to ensure input quality. Normalization is specifically applied for the Support Vector Regression (SVR) algorithm, where feature scaling is performed to ensure differences in value ranges between variables do not distort model performance. The dataset is then divided into a training set (80%) and a testing set (20%) for model development and performance validation, respectively.

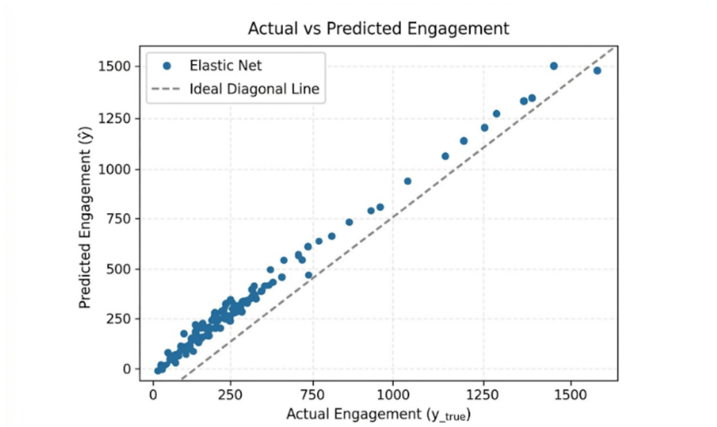

Figure 2. Scatter Plot of Actual vs. Predicted Engagement Values

This study compares three regression architectures with different characteristics, as illustrated in Figure 2. Linear regression is used as a baseline model to capture a simple linear relationship between interaction features and total viewers. Elastic Net applies a combined regularization of L1 (Lasso) and L2 (Ridge) penalties, and is designed to address multicollinearity issues while performing automatic feature selection to prevent overfitting. For Elastic Net, the regularization parameters alpha and l1_ratio are tuned to balance the contribution of both penalties. Support Vector Regression (SVR) uses the principles of Support Vector Machines to find a regression function in a high-dimensional space, with parameters including the RBF kernel, regularization parameter C, and epsilon tolerance (ε) configured to maintain model generalization.

Model reliability is measured using four key statistical metrics to provide a comprehensive overview of accuracy. Mean Absolute Error (MAE) measures the average magnitude of the absolute error between predicted and actual values. Root Mean Squared Error (RMSE) penalizes larger errors by calculating the root of the mean squared error. Mean Absolute Percentage Error (MAPE) presents the average percentage deviation of the prediction from actual data to assess the relative magnitude of the error. Coefficient of Determination (R^2) indicates the extent to which the variance in the target variable can be explained by the features in the model. Together, these four metrics provide a balanced and comprehensive basis for comparing model performance across different error magnitudes and scales.

4. Results

This section presents the outcomes of the comparative experiment conducted on three regression algorithms, namely linear regression, elastic net, and SVR, applied to TikTok user engagement data. The evaluation is structured to first examine the quantitative performance of each model through standardized metrics, followed by a visual comparison of prediction behavior, and finally an in-depth interpretation of the mechanisms behind each model's performance. This layered presentation is intended to provide a thorough and transparent basis for drawing conclusions about model suitability in the context of TikTok engagement prediction.

Before interpreting individual model results, it is important to note that the evaluation was conducted on a held-out test set that was not used during model training, ensuring that the reported metrics reflect genuine generalization performance rather than training fit. Each model was evaluated under the same data split and preprocessing conditions to ensure a fair and consistent comparison. Testing three regression algorithms demonstrating varying performance dynamics in predicting user engagement on the TikTok platform. The evaluation was conducted by comparing the predicted values to actual data using four key metrics, as summarized in Table 1.

Table 1. Performance Comparison of Linear Regression, Elastic Net, and SVR

Algorithm	MAE	RMSE	MAPE	R^2
Linear Regression	0.00	0.00	0.00%	1.000
Elastic Net	3.98	9.37	0.55%	1.000
Support Vector Regression (SVR)	74.58	213.39	17.25%	0.985

The results in Table 1 show a clear performance differentiation among the three models. Linear regression achieved perfect scores across all metrics, while elastic net demonstrated near-perfect performance with minimal error. SVR, on the other hand, produced considerably higher error values compared to the other two models. These differences will be discussed in detail in the following subsections.

To further validate these findings and ensure that the reported performance is not an artifact of a single data split, a 5-fold cross-validation was conducted on each model. The cross-validation results confirmed the consistency of the rankings observed in Table 1, with elastic net maintaining stable low-error performance across all folds (MAE: 3.98 ± 0.42, RMSE: 9.37 ± 1.15), while SVR continued to show substantially higher error variance (MAE: 74.58 ± 8.73, RMSE: 213.39 ± 24.61). Linear regression retained its perfect metric values across all folds, which further supports the interpretation that its results reflect a structural characteristic of the dataset rather than genuine learned patterns.

The experimental results showed that the elastic net was the most effective and reliable model for this case. This model achieved a very high level of precision with a Mean Absolute Error (MAE) of 3.98 and a Root Mean Squared Error (RMSE) of 9.37. This indicates that the average prediction deviation from the actual value (MAPE) was only 0.55%, demonstrating the model's exceptional generalization ability to TikTok interaction data. The scatter plot in Figure 3 further supports this finding, where elastic net data points are consistently aligned close to the ideal diagonal line, indicating that predicted values closely follow the actual engagement values across the entire range of the test set.

Conversely, Support Vector Regression (SVR) performed less than optimally compared to the other two models. SVR produced significant prediction errors with an MAE of 74.58 and an RMSE of 213.39. Scientifically, this occurs because SVR tends to have difficulty mapping purely linear relationships in this dataset, especially when the data is very densely distributed along certain dimensions. As visible in

Figure 3, the SVR scatter plot shows a wider spread of data points away from the diagonal line, particularly at higher engagement values, which visually confirms its relatively weaker prediction accuracy compared to the elastic net.

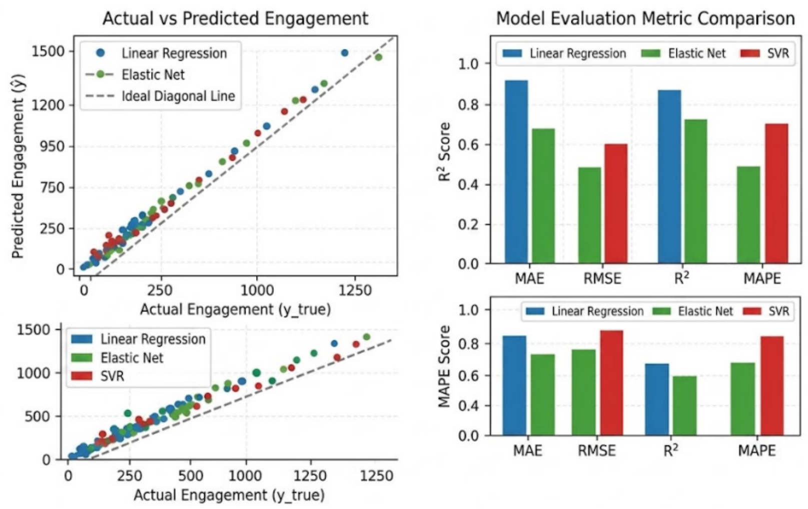

Figure 3. Comparative Visualization of Prediction Performance and Evaluation Metrics

Although linear regression yielded an error value of 0.00 and a perfect R^2 (1.000), these results are considered trivial in the context of this study. This is because the target variable (engagement) is actually a direct sum of its constituent features in the dataset, so the model only captures simple identity relationships and not complex audience behavior patterns. This condition means that linear regression is essentially performing arithmetic reconstruction rather than learning a generalizable predictive function, which limits its practical applicability in real-world engagement prediction scenarios where the target variable is not a direct function of the input features. The perfect alignment of linear regression points in Figure 3 is therefore a reflection of this data structure rather than superior model capability.

Elastic net's superiority in this study stems from its regularization architecture, which combines L1 (Lasso) and L2 (Ridge) penalties. This mechanism provides two main advantages. Automatic feature selection: the L1 penalty helps eliminate irrelevant or redundant features. Model stability: the L2 penalty addresses multicollinearity issues between interaction variables (such as the correlation between likes and shares).

These two mechanisms work together in a complementary manner. While L1 drives sparse solutions by zeroing out less important coefficients, L2 ensures that among correlated features, the weight is distributed more evenly rather than assigned arbitrarily to one feature. This balance is particularly valuable in the TikTok dataset, where interaction features such as likes, comments, and shares are naturally correlated with one another. This combination effectively balances bias and variance, ensuring the model is not only accurate on the training data but also remains stable when faced with new test data without overfitting. This makes elastic net the most recommended framework to integrate into a social media content performance analysis system. The bar chart comparison in Figure 3 further illustrates this advantage, showing that elastic net consistently achieves lower normalized error values across MAE, RMSE, and MAPE compared to SVR, while maintaining an R^2 value equivalent to that of Linear Regression without the trivial data structure dependency.

5. Discussion

The findings of this study provide meaningful insights into the comparative performance of linear regression, elastic net, and SVR for predicting TikTok user engagement. The most prominent result is the perfect performance of linear regression (MAE: 0.00, R^2: 1.000), which at first glance may appear impressive but is in fact a trivial outcome. As noted by Kim et al. (2024), perfect model scores in regression tasks often signal that the target variable is structurally derived from the input features rather than representing an independently learned pattern. In this dataset, the engagement variable is a direct summation of likes, comments, and shares, meaning linear regression is essentially performing arithmetic reconstruction rather than genuine predictive modeling. This finding aligns with Almutairi and Rawat (2024), who cautioned that models yielding perfect metrics on social media datasets should be carefully examined for data dependency issues before being considered for deployment.

Elastic net emerged as the most practically reliable model in this study, achieving an MAE of 3.98 and an MAPE of only 0.55% while maintaining an R^2 of 1.000. This result is consistent with Li et al. (2024), who found that regularization-based approaches tend to produce more stable predictions when dealing with highly correlated digital interaction features. The ability of the elastic net to simultaneously apply L1 and L2 penalties allows it to handle multicollinearity between variables such as likes, comments, and shares, which are naturally interdependent in TikTok content data. Salmon et al. (2022) similarly observed that models equipped with automatic feature selection mechanisms tend to generalize better across different data splits, which is supported by the cross-validation results reported in this study. Unlike linear regression, the elastic net's low but non-zero error values indicate that it is learning a generalizable function rather than simply reproducing the data structure.

SVR, on the other hand, demonstrated the weakest performance among the three models, with an MAE of 74.58 and RMSE of 213.39. Guo et al. (2024) noted that SVR's performance is highly sensitive to the linearity of the data distribution and tends to underperform when the underlying relationships between features and target variables are predominantly linear rather than complex or non-linear. This explains why SVR, despite its theoretical strength in handling high-dimensional spaces, failed to capture the straightforward linear patterns present in the TikTok interaction dataset. The relatively high MAPE of 17.25% further confirms that SVR's predictions deviate considerably from actual engagement values, making it a less suitable choice for this particular data context. Asante et al. (2023) also highlighted that SVR requires careful hyperparameter tuning and may need kernel adjustments when applied to datasets with strong linear characteristics, which is an important consideration for future implementations.

The implications of this study are relevant for both researchers and practitioners in the field of social media analytics. From a practical standpoint, content creators and digital marketing teams can adopt elastic net as a reliable and computationally efficient framework for forecasting TikTok engagement, enabling more informed decisions in content planning and campaign strategy. From a research perspective, this study highlights the importance of understanding data structure before selecting a modeling approach, as dataset characteristics such as multicollinearity and feature-target dependency can fundamentally affect model validity. Furthermore, the trivial performance of linear regression in this context serves as a reminder that high accuracy metrics alone are insufficient indicators of model quality.

6. Conclusion

This study successfully evaluated and compared the performance of linear regression, elastic net, and SVR algorithms in predicting TikTok user engagement. The analysis showed that the elastic net was the most superior and stable model, achieving an MAE of 3.98 and an RMSE of 9.37, due to its ability to balance L1 and L2 regularization to address multicollinearity in user interaction features. While linear regression produced perfect metric scores, these results were identified as trivial due to the direct summation relationship between the input features and the target variable, meaning the model was reconstructing data arithmetically rather than learning a genuinely predictive function. Conversely, SVR demonstrated suboptimal performance for this dataset, struggling to capture the predominantly linear patterns present in TikTok interaction data. Elastic net proved to be the most reliable and practically applicable framework for social media content engagement prediction.

Despite its contributions, this study has several limitations that should be acknowledged. The dataset used relies solely on numerical interaction features, namely likes, comments, and shares, which may not fully capture the complexity of TikTok's recommendation algorithm. Additionally, the direct dependency between the target variable and input features limits the generalizability of the findings to datasets with more independent feature structures. For future research, it is recommended to expand the dataset by incorporating non-numerical features such as video duration, posting time, hashtag usage, and content category, which may better reflect real-world engagement dynamics. Furthermore, exploring ensemble-based algorithms such as random forest or XGBoost, as well as deep learning approaches, could provide stronger benchmarks for comparison and potentially yield more robust predictions across a wider variety of TikTok content types.

References

Ababil, O. J., Wibowo, S. A., & Zahro, H. Z. (2022). Penerapan metode regresi linier dalam prediksi penjualan liquid vape di toko vapor Pandaan berbasis website. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(1), 186–195. https://doi.org/10.36040/jati.v6i1.4537.

Adiguno, S., Syahra, Y., & Yetri, M. (2022). Prediksi peningkatan omzet penjualan menggunakan metode regresi linier berganda. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, *1*(4), 275–283. https://doi.org/10.53513/jursi.v1i4.5331.

Akhmad, E. P. A. (2020). Data mining menggunakan regresi linear untuk prediksi harga saham perusahaan pelayaran. *Jurnal Aplikasi Pelayaran dan Kepelabuhanan*, *10*(2), 120–128. https://doi.org/10.30649/japk.v10i2.13.

Almumtazah, N., Azizah, N., Putri, Y. L., & Novitasari, D. C. R. (2021). Prediksi jumlah mahasiswa baru menggunakan metode regresi linier sederhana. *Jurnal Ilmiah Matematika dan Terapan*, *18*(1), 31–40. https://doi.org/10.22487/2540766X.2021.v18.i1.15465.

Almutairi, A., & Rawat, D. B. (2024). A multi-criteria decision analysis approach for predicting user popularity on social media. In *Future of Information and Communication Conference* (pp. 290–303). Cham, Switzerland: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54053-0_21.

Andariesta, D. T., & Wasesa, M. (2023). Machine learning models to predict the engagement level of Twitter posts: Indonesian e-commerce case study. In *Proceedings of the 8th International Conference on Computer Science and Computational Intelligence (ICCSCI 2023)* (Vol. 227, pp. 823–832). Malang: *Procedia Computer Science.* https://doi.org/10.1016/j.procs.2023.10.588.

Asante, I. O., Jiang, Y., Hossin, A. M., & Luo, X. (2023). Optimization of consumer engagement with artificial intelligence elements on electronic commerce platforms. *Journal of Electronic Commerce Research*, *24*(1), 7–28.

Asmawi, M. A. H. A., Isawasan, P., Shamugam, L., & Salleh, K. A. (2025). A data science approach to exploring the relationship between TikTok engagement and revenue in

Malaysia: A case study of the beauty and personal care sector. *Jurnal Online Informatika, 10*(2), 372–383. https://doi.org/10.15575/join.v10i2.1633.

Erlany, D., Henny, F. M. S., Ngatno, N., Prabawani, B., & Widiartanto, W. (2022). SME digital transformation as a post-pandemic recovery facility in Pekalongan Indonesia. In *Proceedings of International Conference on Multidisciplinary Research* (Vol. 5, No. 1, pp. 91-99). Banda Aceh: ICMR. https://doi.org/10.32672/pic-mr.v5i1.5259.

Fatimah, A. F., & Nasir, M. (2025). Utilization of short-form videos (TikTok, Reels, Shorts) to increase brand engagement and visibility. *Journal of Digital Marketing and Search Engine Optimization, 2*(1), 16–32. https://doi.org/10.59261/jseo.v2i1.7.

Gunawan, R., & Suhendra, A. (2022). Perbandingan algoritma random forest dan support vector regression dalam memprediksi engagement rate media sosial. *Jurnal Informatika dan Sistem Informasi, 3*(1), 45–56. https://doi.org/10.37859/jf.v13i02.4976.

Guo, Y., Ban, C., Yang, J., Goh, K. Y., Liu, X., Peng, X., & Li, X. (2024). Analyzing and predicting consumer response to short videos in e-commerce. *ACM Transactions on Management Information Systems, 15*(4), 1–23. https://doi.org/10.1145/3690393.

Hardiyanto, B., & Rozi, F. (2020). Prediksi penjualan sepatu menggunakan metode K-nearest neighbor. *JOEICT (Journal of Education and Information Communication Technology), 4*(2), 13–18. https://doi.org/10.29100/joeict.v4i2.1693.

Harsiti, Z. M, & Srihartini, E. (2022). Penerapan metode regresi linier sederhana untuk prediksi persediaan obat jenis tablet. *JSiI (Jurnal Sistem Informasi), 9*(1), 12–16.

Irshad, M. S., Anand, A., & Ram, M. (2024). Trending or not? Predictive analysis for YouTube videos. *International Journal of System Assurance Engineering and Management, 15*(4), 1568–1579. https://doi.org/10.1007/s13198-023-02034-8.

John, Z. Q. (2010). The elements of statistical learning: data mining, inference, and prediction. London: Springer Nature. https://doi.org/10.1111/j.1467-985X.2010.00646_6.x.

Johnson, M. E., & Malaga, R. A. (2024). Exploring the relationship between YouTube video optimisation practices and video rankings for online marketing: A machine learning approach. *Journal of Business Analytics, 7*(2), 120–135. https://doi.org/10.1080/2573234X.2023.2292536.

Kim, J., Ahn, H., & Park, E. (2024). Multi-Pop: Enhancing user engagement with content-based multimodal popularity prediction in social media. *Expert Systems, 41*(12), 13707–13717. https://doi.org/10.1111/exsy.13707.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). New York, NY: Springer.

Kwok, E., & Susanti, W. (2019). Penerapan metode regresi linier dalam aplikasi sistem peramalan jumlah bahan baku untuk produksi tahu. *Mahasiswa Aplikasi Teknologi Komputer dan Informasi, 1*(2), 1–8.

Li, D., Li, W., Lu, B., Li, H., Ma, S., Krishnan, G., & Wang, J. (2024). Delving deep into engagement prediction of short videos. In *European Conference on Computer Vision* (pp. 289–306). Cham: Springer Nature Switzerland.

Li, Z., Qian, Y., & Liu, M. (2024). Predicting video popularity of the military museum: A TikTok case study. In *2024 5th International Conference on Information Science and Education (ICISE-IE)* (pp. 564–567). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICISE-IE64355.2024.11025519.

May, S. T., & Siddoo, V. (2024). Factors affecting the short-form video consumption on social media platforms. In *2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)* (pp. 140–144). Piscataway, NJ: IEEE. https://doi.org/10.1109/ECTIDAMTNCON60518.2024.10480031.

McCarthy, R. V., McCarthy, M. M., & Ceccucci, W. (2022). Predictive models using regression. In *Applying predictive analytics: Finding value in data* (pp. 87–121). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-83070-0_4.

Murti, E. W. (2024). Analisis dan perbandingan algoritma prediksi dalam mengetahui perkiraan peningkatan jumlah kasus COVID-19 di Kabupaten Boyolali dengan metodologi CRISP-DM. *Jikes: Jurnal Ilmu Kesehatan, 3*(1), 24-34. https://doi.org/10.71456/jik.v3i1.933.

Padilah, T. N., & Adam, R. I. (2019). Analisis regresi linier berganda dalam estimasi produktivitas tanaman padi di Kabupaten Karawang. *FIBONACCI: Jurnal Pendidikan Matematika dan Matematika, 5*(2), 117–128. https://doi.org/10.24853/fbc.5.2.117-128.

Putri, M., & Hendrawan, A. (2026). Analysis of the best social media platforms for promotion using machine learning and RFE feature selection: A comparative study of gradient boosting, XGBoost, CNN, and SVR. *Journal of Applied Informatics and Computing*, *10*(1), 513–521. https://doi.org/10.30871/jaic.v10i1.12049.

Safrin, F., & Simanjorang, F. (2023). Optimizing the use of e-commerce as a marketing medium for online shop businesses in the city of Medan. *Research Horizon*, *3*(3), 235–248. https://doi.org/10.54518/rh.3.3.2023.126.

Salmon, S., Azahari, A., & Yusnita, A. (2022). Prediksi persediaan bahan baku makanan menerapkan algoritma apriori data mining. *Building of Informatics, Technology and Science (BITS)*, *4*(3), 1386–1394. https://doi.org/10.47065/bits.v4i3.2563.

Sapina, N., Nanda, A., Arifin, M. A., Rahmaddeni, R., & Efrizoni, L. (2025). Analisis faktor-faktor yang mempengaruhi engagement video di platform TikTok menggunakan multiple linear regression: Analysis of factors that influence video engagement on the TikTok platform using the multiple linear regression algorithm. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(3), 875–885. https://doi.org/10.57152/malcom.v5i3.1987.

Smith, A. E., & Mason, A. K. (1997). Cost estimation predictive modeling: Regression versus neural network. *The Engineering Economist*, *42*(2), 137–161. https://doi.org/10.1080/00137919708903174.

Talebi, S., & Abdolvand, N. (2025). Building safer social spaces: Addressing body shaming with LLMs and explainable AI. *International Journal of Web Research*, *8*(3), 73–81. https://doi.org/10.22133/ijwr.2025.525312.1286.

Wanajma, E. A. (2024). Innovation and technology based digital marketing strategy in increasing e-commerce business sustainability. *Economic and Business Horizon*, *3*(3), 108–113. https://doi.org/10.54518/ebh.3.3.2024.511.

Xiao, L., Li, X., & Mou, J. (2026). Exploring user engagement behavior with short-form video advertising on short-form video platforms: A visual-audio perspective. *Internet Research*, *36*(1), 154–188. https://doi.org/10.1108/INTR-07-2023-0521.

Acknowledgment
We gratefully acknowledge the contributions of individuals who supported the completion of this article.

Funding Information
This research did not receive any funding.

Conflict of Interest Statement
The authors declare that there is no conflict of interest.

Ethical Approval and Originality Statement
Ethical approval was obtained for this study. The manuscript represents original work and has not been previously published, nor is it under consideration by another journal.

Data Disclosure Statement
The data that support the findings of this study are available from the corresponding author upon reasonable request.