

Enhancing TikTok Account Performance with Data Pattern Identification

Wilem Musu^{*1}, Nadia Lempan², Indra Samsie³

^{*1}Departement of Informatics Engineering, Dipa Makassar University

²Departement of Digital Business, Dipa Makassar University

³Departement of Information System, Dipa Makassar University

e-mail: *1wilem.musu@undipa.ac.id, 2nadialempan08@gmail.com,
3indrasamsie@diipanegara.ac.id

Abstract

TikTok is a social media platform widely used to share information through short videos to achieve goals and interests in business, education, politics, government, and personal existence. Every activity on this platform is recorded and presented privately to each account owner. However, this data has not been utilized optimally to improve account performance. This research aims to offer a data analysis concept that integrates statistical and machine learning approaches to identify data patterns in each user's data collection, enabling the improvement of account performance. The approach utilizes Linear Regression, k-means, and Decision Tree methods. The results obtained show that the concept of identifying data patterns in TikTok account data has successfully developed a predictive model for video posts that can potentially increase total viewership, video plays, and audience engagement. This is achieved through optimizing video components such as captions, text, hashtags, sound genre, and video type. The outcome yielded a classification model that can predict capable component content to enhance account performance.

Keywords— Data Science Techniques, Data Pattern Identification, Tiktok Account Performance, Statistics, Machine Learning

1. INTRODUCTION

Social media is a new form of media that is capable of forming transnational social networks to share information quickly and is enabled by multimedia [1]. In Indonesia, social media has become a means of sharing information and expression for people of all ages, including students, professionals, politicians, and the government [2]. The top four most popular social media platforms in Indonesia, according to rankings from Reportal, are YouTube, Facebook, TikTok, and Instagram. TikTok currently holds the third position, surpassing Instagram, which now occupies the fourth place [3].

TikTok has emerged as a popular social media platform not only among Generation Z, aged 18–25 years but also among those aged over 35 years [4]. Its popularity stems from providing users with a vast space to share ideas, stories, and creative experiences through short videos [5]. It is not surprising that many users utilize this platform to express themselves as content creators [6]. In the business sector, TikTok serves as an effective medium for reaching the market and conducting various business activities such as promotions, branding, and direct product marketing [7], [8]. The platform's presence has contributed to positive growth in economic sectors, particularly in the digital business domain [9]. Moreover, TikTok's utility extends beyond business; it has also found application in the education sector, serving as a medium for learning, teaching, and knowledge-sharing [10]. Additionally, governmental and political entities have leveraged TikTok as a means of disseminating information [11] [12].

TikTok encourages users to expand their social network by increasing followers, reach,

and interaction levels through the sharing of engaging videos. With the growth of their social network, users can reap various benefits such as increased popularity, trust, and business opportunities [13]. TikTok diligently records all user activities within the application, ranging from browsing, viewing, and searching for content, to engaging with posts by commenting, liking, saving, or resharing, as well as creating and posting their content. These activity logs serve as valuable data for TikTok to enhance user experience, facilitated by the TikTok algorithm a content distribution recommendation technique utilized by the platform [14].

Many TikTok users do not fully grasp that TikTok employs a mechanism or algorithm for content distribution and have not utilized the data provided by TikTok to maximize account performance for better achievement of personal or organizational goals. According to J.J. Smith et al., many TikTok users feel confused and dissatisfied with the algorithmic mechanisms implemented on the social media platform TikTok [15]. While there has been significant research conducted on TikTok, there remains a scarcity of studies that examine and analyze the available data on TikTok accounts to enhance performance for various purposes, including business, education, politics, or government [16]. Another challenge is that managing data on social media platforms like TikTok is not easily achievable using traditional qualitative methods [17].

Recognizing the gap between the underutilization and complexity of managing TikTok insights data and the significance of data analysis outcomes for optimizing both personal and institutional businesses, this research was conducted to present an overview of techniques or methods for analyzing TikTok insights data using quantitative methods with a data science approach. The study focuses on three key metrics: Total Viewers, Plays (Number of Plays), and Audience Interaction (Engagement) because these metrics are widely recognized as critical indicators of content performance and user engagement on TikTok. By focusing on these specific metrics, the research aims to provide a comprehensive understanding of how optimizing video components such as captions, text, hashtags, sound genre, and video type can influence these key performance indicators. Moreover, these metrics are pivotal in assessing the overall impact of content strategies on account performance, aligning with the broader goal of enhancing user experience and achieving organizational objectives on the platform. The aim is to provide knowledge and insights for TikTok users across various objectives, enabling them to harness the vast pool of TikTok data to enhance and optimize their businesses.

Several studies have utilized TikTok insights data to optimize account performance, including research conducted by Vitara et al., published in 2023 [18]. This study aims to examine the correlation between viewer interaction levels with promotional video posts and their purchasing decisions regarding clothing products. The findings indicate a significant impact of audience interaction on the purchasing decisions of young consumers. The insights gained from this research can guide content creators in crafting engaging content that elicits audience likes, comments, shares, or saves. While this study shares similarities with the upcoming research in utilizing data derived from audience interaction, the difference lies in the research objectives.

Another research conducted by Cheng et al. in early 2024 [19] aimed to measure TikTok users' sentiment towards news videos. This research utilized data derived from audience interactions on videos containing news content. The results indicated that videos featuring news content garnered a predominantly negative sentiment. This suggests that TikTok viewers still favor dancing or lip-syncing videos. The findings of this research are valuable for content creators seeking to produce videos with the potential to go viral, incorporating messages through dancing and lip-syncing. While this research shares a similarity with the aforementioned study in utilizing audience interaction as a data source and measurement variable, the key difference lies in its qualitative approach to measuring user sentiment, whereas this research adopts a quantitative approach to enhance account performance.

A study conducted by Wusylko et al., published in early 2024 [17], aimed to address challenges in social media data analysis by introducing machine learning-based analysis

techniques in the context of topic modeling, sentiment analysis, and social network analysis for media platform literacy campaigns across different social media platforms. This research shares a common objective with the upcoming study, as both seek to utilize machine learning approaches to tackle the challenges of analyzing big data available on the TikTok platform. However, the forthcoming study focuses on proposing a data science-based approach to enhance account performance by examining video posts on the platform.

2. RESEARCH METHODS

The data used in this research was entirely taken from TikTok video content on the @nadyalempan account, posted from June 2023 to March 2024. During this period, the account posted more than 170 TikTok videos, but not all were used as data samples. Some videos had very high viewership, which could introduce outlier data or noise into the analysis process. The criteria for selecting data samples for the analysis process were videos watched by 200 to 10,000 users, as recorded in TikTok's total viewer insights (from <https://vt.tiktok.com/ZSYCEwnNB/> to <https://vt.tiktok.com/ZSYCEs2fM/>). Data collection was conducted in early April 2024, with a total of 161 TikTok videos used.

2.1. Data Collection Technique

Data for each piece of content is collected using TikTok's insight feature and tabulated into a data table using the Microsoft Excel application. The recorded data variables include posting time, Total Play Time (TPT), Average Watch Time (AWT), percentage of viewers who Watch the Full Video (WV), New Followers (NV), number of plays (Play), engagement metrics such as Likes, Comments, Shares, and Saves, as well as viewer demographics such as Total Viewers (TV), Returning Viewers (RV), New Viewers (NV), Followers (Fol), Non-Followers (NF), Gender, and Age. All of the variables are shown in Table 1.

Table 1. Variables Visualization.

Date	Time	TPT	AWT (s)	WV (%)	NF	Play	Like	Comment	Share	Save	TV	RV	NV	Fol	NF	Male	Female	18-24	25-34	35-44	45-54	55+
31/03/2024	15:34	07:03:04	6.0	3.33	5	3778	204	4	3	10	3310	42	58	4	96	32	67	75	22	1	1	1
30/03/2024	18:47	04:16:12	4.9	4.47	0	2869	242	0	5	17	2320	67	33	3	97	20	79	72	20	3	2	3
27/03/2024	20:49	00:58:30	3.4	6.74	5	923	108	3	1	8	809	80	20	13	87	73	25	74	22	3	0	1
27/03/2024	11:11	00:50:08	7.1	23.81	0	380	37	0	0	0	316	93	7	51	49	54	46	70	23	3	1	2
25/03/2024	18:00	01:13:51	8.1	30.64	0	550	35	0	0	0	414	94	6	48	52	52	48	71	25	2	0	2
24/03/2024	21:48	11:04:14	8.1	16.77	2	4884	167	0	2	4	4393	36	64	3	97	36	61	73	19	3	2	3
22/03/2024	18:48	00:57:12	3.7	5.68	1	911	89	0	0	6	841	82	18	14	86	66	34	74	21	2	1	2
21/03/2024	15:02	00:45:34	3.0	4.86	2	916	64	1	0	3	821	73	27	14	86	63	37	72	22	3	0	3
20/03/2024	08:06	02:12:40	10.9	7.5	2	720	73	0	0	1	630	90	10	47	53	59	40	70	25	2	1	2
19/03/2024	16:45	03:36:16	6.8	8.69	0	1877	89	1	0	4	1700	68	32	9	91	29	70	75	20	2	1	2
19/03/2024	09:54	01:12:35	4.3	11.88	0	997	83	0	0	0	883	76	24	13	87	53	47	70	24	3	1	2
17/03/2024	17:45	07:23:51	8.4	4.29	0	3138	96	2	0	2	2836	75	25	6	94	27	72	73	24	1	1	1

2.2 Data Analysis Technique

The collected data is then validated manually to ensure there are no missing or incorrectly recorded data points and to assess the extent to which the data meets the criteria in to avoid producing outlier data during the analysis process. The 161 data points used were then analyzed using descriptive statistical methods to understand the dataset's characteristics. If any data are found that do not support (outlier) the analysis process, they are eliminated. The elimination process involves calculating the interquartile range (IQR). The next step involves employing statistical techniques to measure the correlation between variables using the linear regression method, ensuring that the variables being measured influence each other. Once it's confirmed that the data to be analyzed significantly influences the analysis, data grouping/clustering is performed using the k-Means algorithm with three variables: Total Viewer, Play, and Audience Interaction. This clustering aims to identify groups of videos, that exhibit both high and low performance. The clustering results are then used as classification labels/classes to identify classification

patterns in the @nadyalempan account's data overlay. The features in this modeling dataset include Caption, Text, Hashtag, Sound Genre, and Video Type. The analysis process utilizes Python programming language tools with the Pandas, NumPy, Scikit-Learn, Matplotlib, and Statsmodels libraries.

3. RESULTS AND DISCUSSION

3.1 Account Overview

The first TikTok video was posted on the @nadyalempan account in February 2020. By the end of March 2024, more than 2,100 TikTok videos had been uploaded, featuring a diverse range of content. The account boasts a total of 17,780 followers, with some videos garnering over 5 million views.

This research utilizes video content as a dataset extracted from posts spanning from June 2023 to March 2024. Figure 1 illustrates the distribution of viewers during this period, segmented by gender (a), age (b), distinction between repeat and first-time viewers of video posts (c), and differentiation between followers and non-followers (d). The average viewership per video post stands at 2078 viewers.

The average results obtained (Figures 1a and 1b) indicate that the gender distribution of viewers watching video content is nearly equal, with women comprising 56% and men 44%. In terms of age demographics, the majority of the audience falls within the 18-24 age range, accounting for 71%, followed by the 25-34 age range at 24%. The audience aged over 34 years constitutes a smaller proportion, approximately 5%. Considering the functioning of the TikTok algorithm in content dissemination upon posting [20], it becomes evident that the videos shared on the @nadyalempan account resemble or align with the content typically posted by female content creators aged between 18 and 24 years. Despite the relatively minor disparity in viewership between female and male audiences, this suggests that the video posts cater to the preferences of female viewers within the 18-24 age bracket.

Figure 1c depicts data regarding the average engagement of video viewers. It shows that 49% of viewers have watched previous video posts within the past year, while 51% are viewers encountering a video post from this account for the first time. The data suggests that first-time viewers have a higher engagement rate compared to those who have watched previous video posts. This implies that the majority of viewers may not be interested in the characteristics of videos posted on the @nadyalempan account. These viewers might have stumbled upon the video on their homepage (FYP) due to TikTok's algorithm. This underscores the opportunity to enhance account performance to cultivate a more loyal viewer base for future video posts.

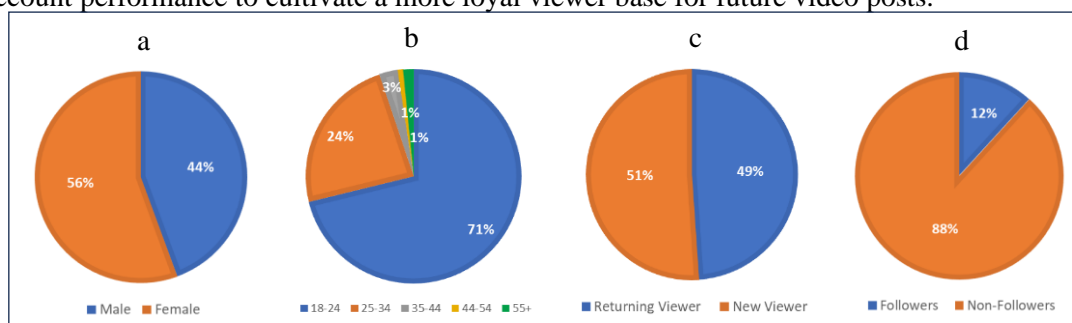


Figure 1. General description of the audience for video posts on the TikTok account @nadyalempan. a) Viewers categorized by gender, b) viewers categorized by age range, c) new viewers versus returning viewers, d) viewers categorized by followership status (followers or non-followers).

Figure 1d indicates that the average video audience presentation consists of 12% from account followers and 88% from non-account followers. This illustrates that the majority of video viewers are individuals who have not yet followed the @nadyalempan account. On one hand, this suggests the effective utilization of the TikTok algorithm by the account, enabling it to reach a

wide audience beyond its followers. However, despite the account's ability to attract a large number of viewers, it has not successfully converted this broader reach into loyal viewership or followers. This aligns with the earlier data, where the proportion of new viewers surpasses that of repeat viewers of video posts on the account. By comparing the presentation of repeat viewers with that of followers, it can be inferred that a significant portion of loyal video viewers have yet to follow the @nadyalempa account.

To enhance account performance as delineated earlier, this research employs statistical and machine learning methodologies to uncover insights in the form of models or patterns for enhancing account performance.

3.2 Data Analysis

The sample data utilized in this research comprised 161 videos. To fulfill the research objectives, three dataset variables were identified: Total Viewers (representing the total number of accounts reached and watching video posts), Play Count (the number of times videos were played by reached accounts), and Interaction, encompassing likes, comments, shares, and saves provided by viewers after watching each video post. Following the determination of these variables, data analysis was conducted employing a statistical approach to ensure the quality and relevance of the data, while also identifying and addressing any outliers.

Data analysis includes the examination of data distribution and dispersion [21]. Data distribution analysis was conducted to determine the frequency of data occurrence across the measured variables of each video post. Based on the computed measures of central tendency, the mean value of the three variables exceeds the median value. Figure 2 depicts a positively skewed normal distribution graph. This indicates that the Total Viewer, Play, and Viewer Interaction data for each video post tend to cluster towards the right side of the graph. The clustering of data around the mean suggests the presence of some data points with high values, potentially leading to outlier data in subsequent analysis.

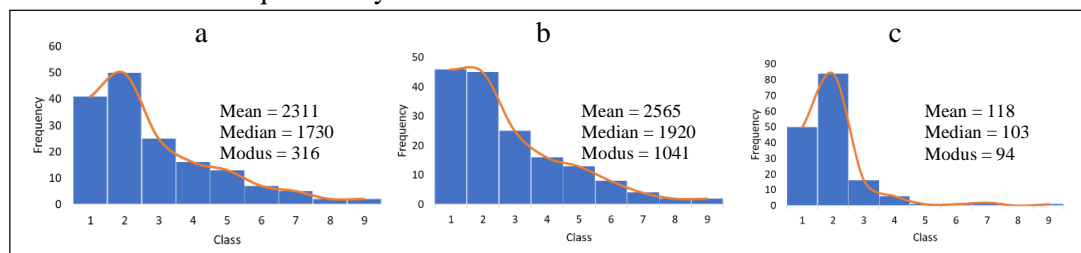


Figure 2. Displays the Data Distribution Graph before Dataset Validation, showcasing: a) Total Viewer Data Distribution, b) Play Data Distribution, and c) Viewer Interaction Data Distribution.

Dispersion analysis is utilized to provide an overview of the proximity or divergence of data distribution from the average value of the variable [22] in each posted video. The results of the data dispersion calculation are depicted in Figure 3 in the form of quartile deviation. Among the 161 videos posted, the Total Viewers ranged from 1000 to 3000 viewers. However, several video posts also garnered around 7000 to 8000 viewers. The analysis reveals the presence of two outlier data points in the Total Viewer variable (Figure 3a). In Figure 3b, it is evident that some video posts have more than one play (Play), resulting in slightly higher values for this variable compared to Total Viewer. The illustration displays two outlier data points in the Play variable. Regarding the Audience Interaction variable, audience response to each video post spans from 70 to 120 interactions. Nonetheless, certain video posts witnessed audience interactions exceeding 600 interactions. Figure 3c illustrates the abundance of outlier data points in the Audience Interaction variable.

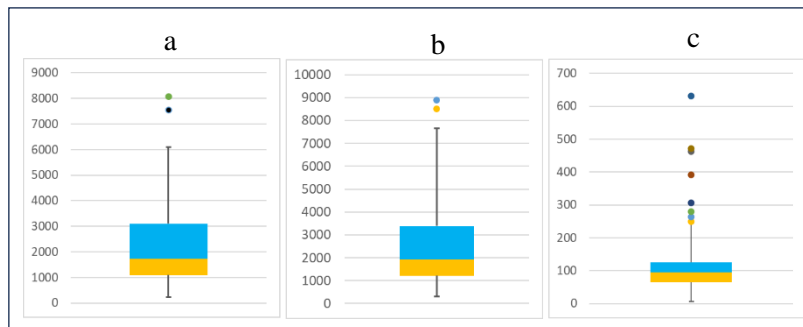


Figure 3. Illustrates the Dataset Quartile Deviation, displaying a) Boxplot Distribution of Total Viewer Data, b) Boxplot Distribution of Play Data, and c) Boxplot Distribution of Viewer Interaction Data.

The analysis of the data conditions outlined above indicates a need for improvement in the collected data before commencing the analysis aimed at identifying patterns of increasing account performance. The technique employed for data enhancement involves eliminating outlier data identified during the dispersion analysis phase. Specifically, two outlier data points were identified in the Total Viewer and Play variables, whereas eight outlier data points were found in the Audience Interaction variable and subsequently removed. Consequently, the dataset underwent a reduction from 161 to 149 entries. Subsequently, the dataset underwent further analysis utilizing frequency distribution techniques, as depicted in Figure 4. The graph illustrating data distribution indicates an enhancement in the overall distribution, although it still exhibits a positive skew. Nevertheless, there is a notable convergence between the average and median values, with the Audience Interaction variable trending toward a symmetrical curve. These two data preparation techniques are paramount for the effective analysis of social media data due to the diverse nature of the data characteristics.

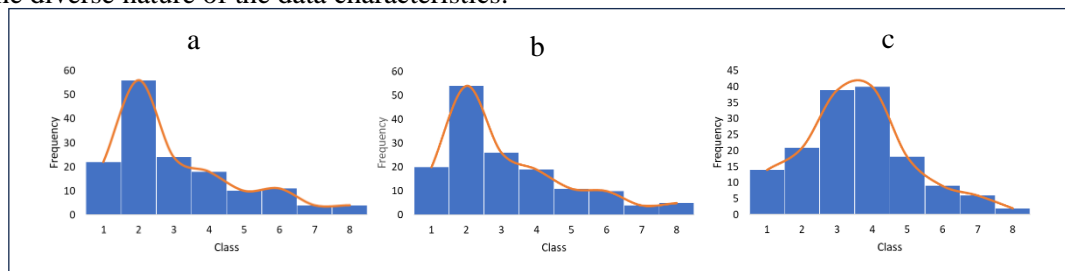


Figure 4. Displays the Data Distribution Graph after Dataset Validation, showcasing: a) Total Viewer Data Distribution, b) Play Data Distribution, and c) Viewer Interaction Data Distribution.

3.3 Data Modeling

To extract insights from data patterns aimed at enhancing the performance of your TikTok account, a series of steps employing statistical and machine learning techniques is undertaken. The initial step involves assessing the correlation levels among variables within the dataset to ascertain the significance of their influence. Subsequently, data is clustered or grouped to identify videos with optimal performance. Finally, the dataset features, including Caption, Text, Hashtag features, Sound Genre, Video Type, and Label, are scrutinized to discern their components. These identification outcomes are then consolidated into a data modeling dataset utilizing the Decision Tree method.

3.3.1 Correlation between Variables

The measurement of the correlation between variables utilizes the linear regression method, where Total Viewers and Play serve as independent variables, while Viewer Interaction acts as the dependent variable, with a significance threshold set at 0.05. The measurement outcomes presented in Table 1, employing the Ordinary Least Squares (OLS) method [23] reveal several crucial findings in this study. Specifically, the correlation level among the three variables

is reflected in regression coefficients, denoted as $\alpha = 0.31$, $\beta_1 = -3.5$, and $\beta_2 = 3.73$. These coefficients constitute components of a mathematical equation, thereby forming a linear regression model capable of predicting the extent of audience interaction, based on the viewership count and the frequency of video post views.

Table 2. Displays the Results of the Correlation Calculation between Variables using Linear Regression.

	coef	std err	t	P > t	[0.025	0.975]
const	0.3101	0.027	11.644	0.000	0.257	0.363
TV	-3.5000	1.484	-2.358	0.020	-6.433	-0.567
PLAY	3.7296	1.471	2.535	0.012	0.822	6.638

Other information obtained from the results of the regression analysis in Table 2 shows that the p-values from the t-test results for all independent variables are smaller than the significance threshold of 0.05. This indicates that the Total Viewer and Play variables significantly influence the Audience Interaction variable. However, the large standard error value (std err) suggests that these two variables have a confidence level below 95% when the resulting model is used to predict other data samples from the population used. These results illustrate that the creation of video content on this account does not yet follow a specific pattern, resulting in significant variation in audience reach and interaction between each video post.

Regression analysis of the data variables for each video post available in the video analysis feature can provide clear knowledge and insights about account characteristics and the relationships between data variables that influence each other. This knowledge can be utilized to develop a predictive model aimed at enhancing TikTok account performance.

3.3.2 Data Clusterization

Based on the results of the regression analysis, it was determined that video posts on the @nadyalempa account elicited highly diverse responses, highlighting the necessity for further modeling steps to pinpoint the optimal model for enhancing account performance. To achieve this, data grouping was undertaken to identify the data subset exhibiting the most favorable performance, utilizing the same dataset employed in the regression analysis.

Before proceeding with the clustering analysis, data normalization was conducted for all variables within the dataset due to variations in value ranges among the Total Viewer, Play, and Viewer Interaction variables. The determination of the number of clusters employed the Elbow technique, which facilitates the optimization of cluster count based on dataset characteristics by calculating the inertia value for each cluster model. As depicted in Figure 5a, a significant inflection in the inertia value is observed at the second point, where the Elbow graph forms a distinct elbow, indicating a notable decrease in the inertia value [24]. This signifies that the optimal number of data clusters is two. Increasing the number of clusters beyond two would yield negligible effects, as the inertia value no longer demonstrates a significant decrease.

Figure 5b shows that the data distribution is segmented into two clusters: the blue data points represent the first cluster, while the yellow data points constitute the second cluster dataset. Upon examining the distribution within each cluster, it becomes apparent that there is significant variability in the data, as evidenced by the considerable distance between the data points and the centroid (depicted by the red cross). Nonetheless, the outcomes of this analysis serve to offer valuable insights into data that exhibits commendable performance, thus serving as a potential model for enhancing account performance.

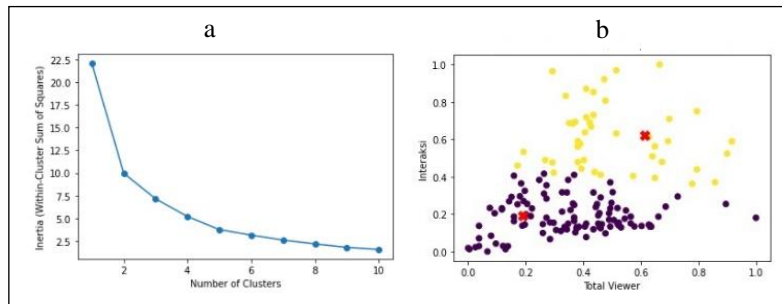


Figure 5. Results of Clustering Analysis: a) Determination of the Number of Clusters using the Elbow Method. b) Data Clusterization Results.

The first cluster comprises 105 data points, while the second cluster comprises 44 data points. Table 3 illustrates that video posts in the second cluster exhibit superior performance compared to those in the first cluster. Regarding the Total Viewer variable, the lowest viewership count for video posts is 2363, with the highest reaching 6100 viewers. In terms of video playback, the minimum number of plays is 2636 times, while the maximum is 6582 times. Additionally, Audience Interaction also demonstrates commendable performance, despite the first cluster having a larger maximum value for audience interactions. This suggests that there are video posts in the first cluster with low reach and video playback but high audience interaction.

Table 3. Minimum, Maximum, and Mean Values for Each Data Variable in Each Cluster.

Cluster	Number of Data	Variable								
		Total Viewer			Play			Engagement		
		Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
1	105	241	2678	1341	312	2967	1502	6	245	90
2	44	2363	6100	3836	2636	6582	4204	47	225	126

3.3.3 Classification Models

Data modeling employs a classification method utilizing the Decision Tree method. The dataset's labels/classes stem from data clustering results, grouping sample data into two categories: high and low-performing videos. Subsequently, each feature's values are identified. This includes analyzing caption, text, and hashtag features to determine whether the sample video contains these elements. Additionally, the music/sound genre feature is evaluated to ascertain if the video employs pop, hip hop/rap, electronic/remix, or alternative music/sounds. Video types are categorized as Dance (D), Lip-syncing (LS), Storytelling Video (SV), and Photo Clip (PC).

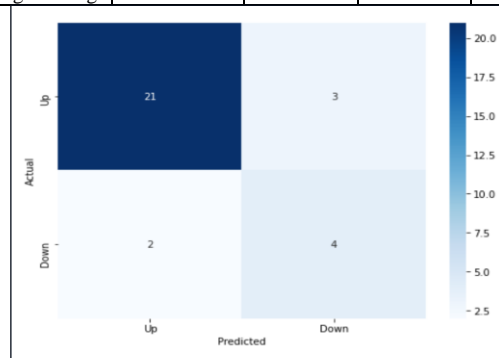
For classification modeling using the C4.5 algorithm, the dataset is split into 80% (109) training data and 20% (40) testing data. The confusion matrix in Figure 6 indicates that out of 24 actual high-performance data, 21 were correctly predicted, while 3 were predicted incorrectly. Among the 6 low-performing data, 4 were correctly predicted, and 2 were predicted incorrectly. The statement explains that the accuracy score of the model in making predictions is 83.33%.

Table 4 illustrates that the resulting model exhibits an 83% confidence level in enhancing video posting performance on the @nadyalempa account. This is evidenced by successfully identifying high-performing videos (91%), while the identification rate for low-performing videos stands at 57%. Notably, 43% of high-performing videos may be misidentified as low-performing ones. To bolster the model's accuracy in predicting new data, a recall of prior identification results was conducted, revealing an 88% ability to identify high-performance videos and a 67% ability to identify low-performance.

The difference between these results was then subjected to an f1-score test to harmonize precision and recall values, yielding an 89% prediction accuracy for high-performance videos and 62% for low-performance ones. The higher predictive accuracy for high-performing videos compared to low-performing ones suggests that the resulting model can effectively enhance TikTok account performance, aligning with the research objectives.

Table 4. Test Results of the Classification Model.

	precision	recall	f1-score	support
Up	0.91	0.88	0.89	24
Down	0.57	0.67	0.62	6
accuracy			0.83	30
macro avg	0.74	0.77	0.75	30
weighted avg	0.84	0.83	0.84	30

*Figure 6. Displays the Heatmap of the Confusion Matrix illustrating the Model Testing Results.*

The data modeling results for the TikTok account @nadyalempan, depicted in Figure 7, reveal three patterns conducive to enhancing account performance and attracting viewers with increased engagement. Firstly, Lip-Syncing videos do not necessitate text and hashtags and can encompass various sound genres, provided the sound aligns with the talent's character and addresses contemporary issues, offering solutions through captions. Adding hashtags and text diminishes the video's efficacy. Here, sound serves as the primary narrative element, necessitating talent comprehension of lyrics and music. Secondly, Storytelling videos require captions and text without hashtags, with captions elucidating the storyline and text reinforcing dialogue for audience comprehension. Like the first pattern, sound genre flexibility is recommended, focusing on viral sounds. Lastly, Photo Clip (PC) videos, while less effective in improving account performance, offer variety. These videos require electronic/remix genre sound, viral captions, and structured photo messages complemented by caption explanations. To enhance appeal, photo clip arrangement and display should synchronize with the sound's rhythm, incorporating transitions and popular pause styles.

From the results of identifying this data pattern, it was found that this type of dance video was not very popular among the viewers of this account. However, if you wish to diversify your posts without necessarily aiming to enhance account performance, you can utilize this pattern, including the video components used, such as employing pop genre sounds, without captions and hashtags. Another finding from this research is that the use of hashtags does not significantly impact account performance. The fact that hashtags are at the root of the Decision Tree model indicates their high entropy value, making them a distinguishing feature for class separation in the dataset [25].

The application of data science techniques, including statistical and machine learning approaches, to uncover insights within social media account data, particularly on platforms like TikTok, has demonstrated its effectiveness in providing a comprehensive overview of account performance and other critical aspects essential for optimizing both individual and organizational account management. It is anticipated that the outcomes of this research can serve as a framework for identifying data patterns in accounts created for specific purposes across various domains such as business, education, government, and politics. Understanding these data patterns can effectively streamline account management to align with organizational goals and targets.

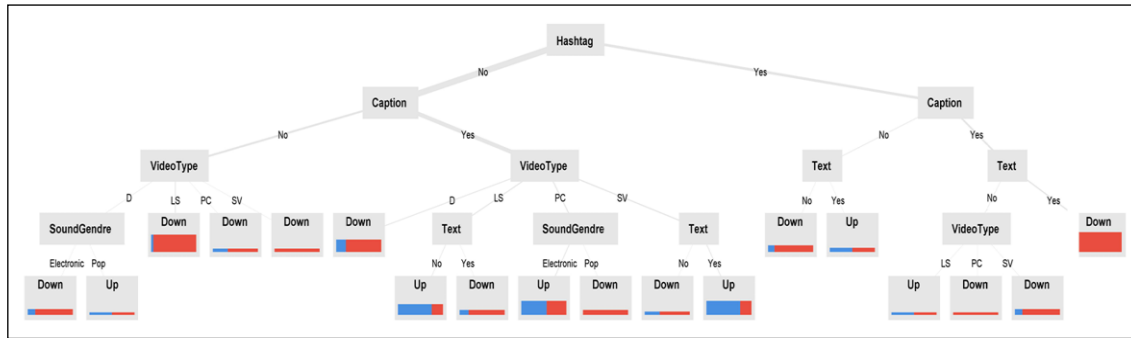


Figure 7. Decision Tree Classification Model for TikTok Video Components

4. CONCLUSION

Based on the results and discussion, several important conclusions can be drawn regarding the identification of TikTok account data patterns as follows:

1. TikTok provides facilities for each user to access data related to video posts, enabling users to leverage this data to achieve their maximum account usage goals.
2. The data available on TikTok accounts contains valuable knowledge and insights for every user. The application of data science techniques involving statistical methods and machine learning is the appropriate approach to uncover and extract insights from the data provided by TikTok.
3. The concept of identifying data patterns using a data science approach that combines Regression, k-means, and Decision Tree algorithms has been demonstrated to effectively reveal insights that can optimize and enhance the performance of TikTok accounts.
4. The dataset available on TikTok holds significant potential for a wide range of knowledge and insights, which has yet to be fully utilized by most users. The concept of identifying data patterns, as demonstrated in this research, can be further applied to discover additional patterns that align with the specific needs and objectives of individual users.

5. RECOMMENDATION

Through the study and analysis conducted in this research using TikTok Insight data, certain significant findings have emerged, highlighting areas that merit deeper exploration through a machine learning framework. Specifically, there is a compelling need to delve further into the classification of sound genres and video types to gain deeper insights into their respective impacts on audience interaction.

The research has revealed initial insights into how sound genres and video types influence viewer engagement on TikTok. However, to comprehensively understand these dynamics and to enhance predictive models, a more nuanced analysis is required. By employing machine learning techniques, such as classification algorithms, it becomes possible to categorize and analyze the intricate relationships between different sound genres and video types, and their respective effects on audience interaction metrics like viewer retention, likes, and comments.

Furthermore, a deeper exploration into these variables could uncover patterns and trends that are crucial for content creators and marketers on TikTok. By leveraging machine learning approaches, this research aims to provide actionable insights that can optimize content strategies, enhance audience engagement, and ultimately improve the effectiveness of TikTok campaigns.

In conclusion, expanding the analysis to include machine learning methodologies will not only deepen our understanding of how sound genres and video types influence audience interaction. But, also smooth the way for more informed decision-making in digital content creation and marketing strategies on TikTok.

6. FUTURE WORKS

By grasping the concept of identifying data patterns to formulate a model aimed at enhancing account performance, subsequent research endeavors will delve into employing this concept to analyze the key elements of videos and sounds that need optimization. This analysis aims to generate a substantial increase in the number of followers for each TikTok video uploaded.

7. ACKNOWLEDGEMENTS

We would like to express our gratitude for the support provided by Dipa Makassar University and the Dipanegara Foundation in funding this research. We recognize that without this support, the progress of our research would not have been as seamless.

DAFTAR PUSTAKA

- [1] F. F. Noorikhsan, H. Ramdhani, B. C. Sirait, and N. Khoerunisa, "Dinamika Internet, Media Sosial, dan Politik di Era Kontemporer: Tinjauan Relasi Negara-Masyarakat," *J. Polit. Issues*, vol. 5, no. 1, pp. 95–109, Jul. 2023, doi: 10.33019/jpi.v5i1.131.
- [2] N. G. Antaguna and A. A. S. L. Dewi, "Pembatasan Kebebasan Berpendapat dan Berekspresi di Sosial Media Berdasarkan Peraturan Perundang-Undangan Nomor 19 Tahun 2016 tentang Perubahan atas Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (ITE)," *Kertha Wicaksana*, vol. 17, no. 2, pp. 138–146, 2023.
- [3] G. Susanto, R. Evelyn, D. Leo, H. Felicio, U. P. Harapan, and M. Abstract, "Pemanfaatan Algoritma Tiktok Dan Instagram Untuk Meningkatkan Brand Awareness," *J. Ilm. Wahana Pendidik.*, vol. 9, no. 22, pp. 483–491, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.10115513>.
- [4] D. H. Najah, A. A. Putra, and N. Aiyuda, "Kecenderungan Narsistik dengan Intensitas Penggunaan Aplikasi Tiktok pada Mahasiswa," *Psychopolytan J. Psikol.*, vol. 5, no. 1, pp. 1–7, Aug. 2021, doi: 10.36341/psi.v5i1.1914.
- [5] A. Aldila Safitri, A. Rahmadhany, and I. Irwansyah, "Penerapan Teori Penetrasi Sosial pada Media Sosial: Pengaruh Pengungkapan Jati Diri melalui TikTok terhadap Penilaian Sosial," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 3, no. 1, pp. 1–9, Jan. 2021, doi: 10.47233/jteksis.v3i1.180.
- [6] J. A. Yani, R. R. Selatan, and Y. Indonesia, "Intensitas Mengakses Aplikasi TikTok dan Pengaruhnya terhadap Komunikasi Interpersonal Remaja," vol. 6, no. 01, pp. 13–24, 2023.
- [7] L. Fawahan and I. M. Purnasari, "Membangun Branding Produk UMKM Melalui Media Sosial Tiktok," *Al-Kharaj J. Ekon. Keuang. Bisnis Syariah*, vol. 4, no. 2, pp. 280–289, Nov. 2021, doi: 10.47467/alkharaj.v4i2.627.
- [8] L. Saffira and G. Rahmawan, "Perilaku Penggunaan Tik Tok Sebagai Media Promosi Pebisnis," *Ekon. Keuangan, Investasi dan Syariah*, vol. 3, no. 4, pp. 717–723, May 2022, doi: 10.47065/ekuitas.v3i4.1451.
- [9] M. B. Priyono and D. P. Sari, "Dampak Aplikasi Tiktok Dan Tiktok Shop Terhadap

- UMKM Di Indonesia,” *J. Ilm. Wahana Pendidik.*, vol. 9, no. 17, pp. 497–506, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.8315865>.
- [10] A. N. Vidyastuti, M. M. Effendi, and R. Darmayanti, “Aplikasi Tik-Tok: Pengembangan Media Pembelajaran Matematika Materi Barisan dan Deret Untuk Meningkatkan Minat Belajar Siswa SMA,” *J. Math Educ. Nasant. Wahana Publ. Karya Tulis Ilm. di Bid. Pendidik. Mat.*, vol. 8, no. 2, pp. 91–106, Nov. 2022, doi: 10.29407/jmen.v8i2.18267.
- [11] E. Effendi, S. Azura, N. A. Harahap, N. Nurmara, and M. H. P. Pratama, “Dualisme Efek Pemanfaatan Aplikasi Tiktok Dalam Humas Pemerintah,” *Innov. J. Soc. Sci. Res.*, vol. 4, no. 1, pp. 1865–1878, 2024.
- [12] N. P. Kurniawan, “Analisis peran Media Sosial ‘TikTok’ di Era Digital menuju Pesta Politik di Tahun 2024 studi kasus calon Presiden Ganjar Pranowo,” *J. Media*, vol. 1, no. 1, 2023, [Online]. Available: <https://jurnal.mediaakademik.com/index.php/jma/article/view/8%0Ahttps://jurnal.mediaakademik.com/index.php/jma/article/download/8/9>.
- [13] R. Siddik, Roswaty, and Meilin Veronica, “Pengaruh Konten Kreatif, Interaksi Pengguna dan Popularitas Influencer Terhadap Keputusan Pembelian Konsumen Pada Program Afiliiasi TikTok,” *JEMSI (Jurnal Ekon. Manajemen, dan Akuntansi)*, vol. 10, no. 2, pp. 1048–1058, Apr. 2024, doi: 10.35870/jemsi.v10i2.2251.
- [14] Z. Chen and C. Shi, “Analysis of Algorithm Recommendation Mechanism of TikTok,” *Int. J. Educ. Humanit.*, vol. 4, no. 1, pp. 12–14, Aug. 2022, doi: 10.54097/ijeh.v4i1.1152.
- [15] J. J. Smith, L. Jayne, and R. Burke, “Recommender Systems and Algorithmic Hate,” in *Proceedings of the 16th ACM Conference on Recommender Systems*, Sep. 2022, pp. 592–597, doi: 10.1145/3523227.3551480.
- [16] R. Rachmadayanti, F. Andini, V. V. Susilo, and W. Sheilma, “DAMPAK ALGORITMA TIKTOK TERHADAP KONSUMSI MUSIK,” pp. 2069–2076, 2021.
- [17] C. Wusylko *et al.*, “Using machine learning techniques to investigate learner engagement with TikTok media literacy campaigns,” *J. Res. Technol. Educ.*, vol. 56, no. 1, pp. 72–93, Jan. 2024, doi: 10.1080/15391523.2023.2266518.
- [18] V. Vitara and M. Kurniawati, “Pengaruh Social Media Engagement Tiktok Terhadap Impulse Buying Produk Pakaian Pada Dewasa Awal,” *J. Humanipreneur*, vol. 2, no. 2, pp. 15–21, Aug. 2023, doi: 10.53091/hum.v2i2.39.
- [19] Z. Cheng and Y. Li, “Like, Comment, and Share on TikTok: Exploring the Effect of Sentiment and Second-Person View on the User Engagement with TikTok News Videos,” *Soc. Sci. Comput. Rev.*, vol. 42, no. 1, pp. 201–223, Feb. 2024, doi: 10.1177/08944393231178603.
- [20] T. N. Fitria, “Value Engagement of TikTok: A Review of TikTok as Learning Media for Language Learners in Pronunciation Skill,” *EBONY J. English Lang. Teaching, Linguist. Lit.*, vol. 3, no. 2, pp. 91–108, Jul. 2023, doi: 10.37304/ebony.v3i2.9605.
- [21] A. Wahab, A. Syahid, and J. Junaedi, “Penyajian Data Dalam Tabel Distribusi Frekuensi Dan Aplikasinya Pada Ilmu Pendidikan,” *Educ. Learn. J.*, vol. 2, no. 1, p. 40, 2021, doi: 10.33096/eljour.v2i1.91.

-
- [22] R.-D. Hilgers, N. Heussen, and S. Stanzel, *Statistik, deskriptive*, no. 1. 2019.
- [23] N. P. Y. Widiasih, A. Sokarina, and Isnawati, “Socio-Economic and Humanistic Aspects for Paylater System and,” *Socio-Economic Humanist. Asp. Townsh. Ind.*, vol. 1, no. 2, pp. 188–197, 2023.
- [24] N. T. Hartanti, “Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 82–89, Aug. 2020, doi: 10.25077/TEKNOSI.v6i2.2020.82-89.
- [25] M. A. Hasanah, S. Soim, and A. S. Handayani, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
-