Engineering and Technology Journal e-ISSN: 2456-3358
Volume 10 Issue 10 October-2025, Page No.-7279-7284
DOI: 10.47191/etj/v10i10.08, I.F. – 8.482

From Sentiment to Strategy: Emotional Classification of Customer Engagement in Corporate Cosmetic Video Marketing

Siti Monalisa*
Departement of Information Systems, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru-Riau, Indonesia.

ABSTRACT: This study examines various emotions expressed in customer engagement content available in corporate product videos on the TikTok platform. The research identified 301 posts with a total of 39,615 comment data collected from the official TikTok account. The data underwent a preprocessing stage to ensure accurate processing, followed by stemming, classification using TF-IDF, Bag-of-Words visualization, and classification with the NRC EmoLex library to detect emotions present in the dataset. The findings reveal that, in addition to positive and negative comments, various emotions such as Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust, Positive, and Negative were identified. The highest emotion score was Positive (33%), followed by Joy (17%), Anticipation (16%), and Surprise (12%). These results indicate that the content analyzed in this study has a substantial positive impact on the audience, with positive emotions dominating. However, companies should remain attentive to the presence of Anticipation and Negative emotions. Therefore, it is essential to maintain factors that trigger positive emotions, minimize content that elicits negative emotions, and manage sensitive content carefully to avoid generating adverse responses.

KEYWORDS: Bag-of-Words (BoW), EmoLex Library, Emotions, Negative Sentiment, Positive Sentiment, TikTok, and Video Comments

I. INTRODUCTION

The emergence of various communication media today has encouraged companies to leverage multiple channels to interact with customers [1]. The availability of channels such as websites, social media, and video-sharing platforms has become an essential component of corporate marketing strategies. Through these channels, users can directly access video content without downloading it [2] and are also able to comment, reply, and express likes or dislikes on shared video content [3].

Currently, both companies and customers utilize social media to interact with video creators or other viewers. Companies often take advantage of video comments to highlight their product advantages, while customers use social media to share their experiences and seek opinions and feedback from other users before making a purchase decision. However, not all video content has the same effectiveness in engaging customers or influencing consumers' purchase intentions. Several elements influence customers' emotional engagement, including verbal narration, facial expressions, and tone of voice in videos. Therefore, companies must pay attention to these elements when developing content strategies. Moreover, user comments and interactions on videos reflect the level of engagement and perception toward the product. Although content strategy plays a crucial role in marketing, there are still limitations in understanding how video content elements affect customer engagement and purchase intention.

The increasing amount of customer engagement content available in corporate product videos may contain hidden dimensions and deeper emotions that ultimately influence customers' purchasing decisions [1]. Emotional engagement drives deeper interactions, where customers who experience specific emotions tend to comment, like, share video content, and are expected to build an emotional bond with the brand and foster long-term loyalty [4]. In addition, specific emotions exert a stronger influence on decisions than general positive attitudes. Therefore, this study investigates how various emotions are expressed in customer engagement content featured in corporate product videos.

In the context of Indonesia, the cosmetic industry has experienced significant growth over the past decade, driven by increasing consumer awareness of beauty trends, self-care, and social media influence [5]. Data from the Indonesian Ministry of Industry indicates that the cosmetic sector is among the fastest-growing industries in the country, supported by a large millennial and Gen Z population that actively engages with digital platforms for beauty inspiration and product recommendations [6]. The rise of e-commerce platforms and short-video applications such as TikTok has further accelerated this trend, providing brands with opportunities to promote products through creative and interactive content [7]. Consequently, beauty and cosmetic brands in Indonesia have adopted video-based marketing strategies to attract and engage customers, leveraging influencers and brand-generated content to strengthen brand loyalty and drive purchase intentions.

Product videos generally originate from two main types: those created by companies themselves, known as brand-generated content, and those produced by consumers using the product or service, known as user-generated content. This study focuses on brand-generated videos on the TikTok platform. Videos produced by the company are used to predict customer behavioral outcomes. Customer behavior can be measured through video comments by analyzing comments, replies, likes, and other forms of interaction, which ultimately lead to positive and negative recommendations influencing consumers' purchasing decisions [8].

Previous literature has examined how customers respond to these videos through commenting, liking or disliking, and replying [9], [10]. All these actions reflect the growing role of customer engagement on online video platforms, which ultimately assists prospective buyers in making purchasing decisions [11] and, in turn, determines the success or failure of a company's product or brand. Post-video customer behaviors have also increased in the form of likes or dislikes, comments, and replies in various contexts [3], [12], [13]. With the rise of social media usage, such behaviors have become more easily observable by customers An increase in social media engagement has enhanced the visibility of these behaviors to customers [1], [14].

II. RESEARCH METHOD

This study collected customer engagement content from product videos available on the TikTok platform, focusing on all comments in their original form. Comments from selected videos were extracted and organized into a structured dataset in tabular format, consisting of qualitative text entries.

The preprocessing phase involved several steps to ensure data quality. Initially, a cleaning process was conducted to remove elements such as emoticons, hashtags, mentions, excessive spaces, numbers, duplicate entries, and comments with fewer than five words [15]. Subsequently, case folding was applied to convert all text to lowercase, followed by tokenization to split sentences into individual words, normalization to standardize linguistic forms, and stop-word removal to eliminate non-informative terms. Stemming was then performed to reduce words to their root form. These procedures facilitated the removal of irrelevant terms and enabled the identification of significant Bag-of-Words (BoW) features using the Term Frequency–Inverse Document Frequency (TF-IDF) algorithm.

The sentiment labeling stage employed a lexicon-based approach using Python to classify words as Positive or Negative. To ensure accuracy, sentiment assignments were validated by an independent expert. Additionally, TF-IDF was utilized to quantify term significance by evaluating its distribution across documents [16]. The BoW representation was then visualized through a word cloud to depict the most frequent terms.

Finally, emotion classification was performed using the NRC Word-Emotion Association Lexicon (EmoLex) through the R programming language. This resource facilitated the extraction of positive and negative sentiments, along with eight primary emotions: joy, trust, surprise, anticipation, anger, fear, sadness, and disgust [16]. Emotion scores were computed based on the frequency of terms associated with each emotional category. Prior studies have extensively documented the applicability of EmoLex-based sentiment analysis across various research domains [17], [18], [19].

III. RESULT AND DISCUSSION

The dataset for this study comprised 301 TikTok posts collected through a web scraping approach, yielding a total of 39,615 comments from the official TikTok account. Following an initial inspection, a comprehensive data cleaning process was conducted to ensure accuracy and consistency. After cleaning, the dataset was reduced to 23,341 comments. The cleaning process was implemented in two stages: a manual review to remove irrelevant entries and automated filtering using Python scripts to detect and eliminate noisy data, such as empty values or unusable entries.

The text preprocessing stage involved multiple sequential steps. First, case folding was performed to convert all text into lowercase format, ensuring uniformity across the dataset. This was followed by tokenization, in which sentences were segmented into individual tokens. Tokenization was carried out using the Natural Language Toolkit (NLTK) library in Python, which is widely recognized for text processing tasks.

Next, data normalization was applied to correct non-standard language forms and abbreviations commonly found in user-generated content on social media platforms. For example, terms such as "uda" were normalized to "sudah" (already), "yg" to "yang" (which), and "sblm" to "sebelum" (before). Despite these efforts, the normalization process did not fully standardize all informal variations, leaving some inconsistencies in the dataset. The outcomes of tokenization and normalization are presented in Tables 1 and 2.

Following normalization, stopword removal was conducted to eliminate words that lack semantic relevance or do not contribute meaningfully to the analysis. This step aimed to improve the quality of the text representation by discarding high-frequency but low-information terms. Examples of removed stopwords include "kakk," "ini" (this), "live," "sudah" (already), "sebelum" (before), "tidak" (not), "atau" (or), "mau" (want), "nya," and "yang" (which). By removing these terms, the dataset was refined to include only linguistically significant tokens, facilitating more accurate subsequent analyses.

These preprocessing steps collectively ensured that the dataset was prepared for advanced text mining techniques, such as Bag-of-Words (BoW) modeling, Term Frequency–

Inverse Document Frequency (TF-IDF) weighting, sentiment labeling, and emotion classification, which are discussed in the following sections.

Table 1. Sample Of Tokenized Data

No	Tokenizing
1	["kakk", "live", "kapann", "uda", "nungguin", "ini"]
2	["dh", "tau", "komenku", "sblum", "live"]
3	["yg", "mau", "beli", "scincare", "cammile", "baru", "pakai", "dikit", "dimurahin"]
4	["aaa", "ketinggalan", "live", "camille", "habis", "ketiduran"]
5	["kak", "mau", "tanya", "masker", "camille", "yang", "varian", "apa", "ya", "untuk", "komedo", "mendem"]
...	...
23.341	["nanti", "mlm, "ad", "ap","ini", "kaka"]

Table 1. Sample Of Normalized Data

No	Tokenizing
1	["kakk", "live", "kapan", "sudah", "nungguin", "ini"]
2	["sudah", "tahu", "komenku", "sebelum", "live"]
3	["yang", "mau", "beli", "skincare", "cammile", "baru", "pakai", "dikit", "dimurahin"]
4	["aaa", "ketinggalan", "live", "camille", "habis", "ketiduran"]
5	["kak", "mau", "tanya", "masker", "camille", "yang", "varian", "apa", "ya", "untuk", "komedo", "tersumbat"]
...	...
23.341	["nanti", "mlm, "ada", "apa","nih", "kaka"]

The final stage of text preprocessing was stemming, which serves to reduce word variations by converting tokens to their root forms, thereby simplifying the text representation within the training data. This process was implemented using the **Sastrawi** module in Python. For instance, words such as "masknya" were converted to "mask", "komenku" to "komen", "nungguin" to "tunggu", and "produknya" to "produk".

Following text preprocessing, sentiment labeling was conducted using a **lexicon-based approach** implemented in Python. In this stage, each comment was assigned either a positive or negative label based on its sentiment score. However, the labeling process revealed some comments with null values or classified as neutral sentiment. To address this, manual cleaning was performed to remove comments without positive or negative sentiment values.

Subsequently, the number of comments in each sentiment category was calculated. The results indicated that **14,869 comments** were classified as positive (represented in green), whereas **8,472 comments** were classified as negative (represented in red). **Figure 4.2** illustrates the proportion of sentiment analysis results for the TikTok comment dataset of Camille Beauty Official. Based on the visualization, the highest proportion corresponds to positive sentiment at **63.7%** (green), while negative sentiment accounts for **36.3%** (red).

The classification results indicate that the number of positive comments substantially exceeds the number of negative comments, suggesting that, overall, user responses toward the analyzed TikTok content were predominantly positive. Nonetheless, the presence of a significant proportion of negative comments highlights the need for continued attention.

Following the classification stage, **TF-IDF** was applied to compute term weights and assign numerical values to the extracted terms. Thereafter, a **Bag-of-Words (BoW)** approach was utilized, visualized in the form of a **word cloud,** where the most frequently occurring words appear in larger font sizes. A subset of comments was randomly selected from the original dataset for this visualization. The analysis revealed dominant terms such as "camille", "banget", "dapet", "beli", and "kak", which are significant indicators of purchase intent. The BoW visualization is presented in **Figure 1.**

Figure 1. Bag-of-Words Visualization of Positive Sentiment

In the TikTok comments for Camille Beauty, the two most frequently occurring words associated with positive sentiment (indicated in green) were "thank" and "birthday." In contrast, for negative sentiment (indicated in red), the most frequent words were "perih" and "kalah."

The next stage involved emotion analysis, which was performed in Google Colaboratory using EmoLex (NRC Word-Emotion Association Lexicon) to extract positive and negative sentiments as well as associated emotions, including Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, and Trust (Mohammad & Turney, 2012). Positive sentiment was associated with the emotions Anticipation, Joy, Surprise, Trust, and Positive, while negative sentiment was linked to Anger, Disgust, Fear, Sadness, and Negative.

Figure 2 presents the results of the emotion classification, showing that the categories Positive and Joy recorded the highest scores, followed by Trust and Anticipation, which were also relatively high. This indicates that the data predominantly conveys optimism and trust. Conversely, negative categories such as Anger, Disgust, Fear, and Sadness had lower scores and were therefore not dominant. The emotions Surprise and Negative were also present but not significant.

The emotion analysis revealed ten categories: Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust, Positive, and Negative (Figure 3). The highest-scoring emotion was Positive (33%), followed by Joy (17%), Anticipation (16%), and Surprise (12%).

These findings indicate that Camille Beauty's content has a predominantly positive impact on its audience, with positive emotions being dominant. However, the presence of Anticipation and Negative emotions suggests that the company should continue to foster positive triggers, reduce content that evokes negative responses, and carefully manage sensitive material to prevent adverse reactions.

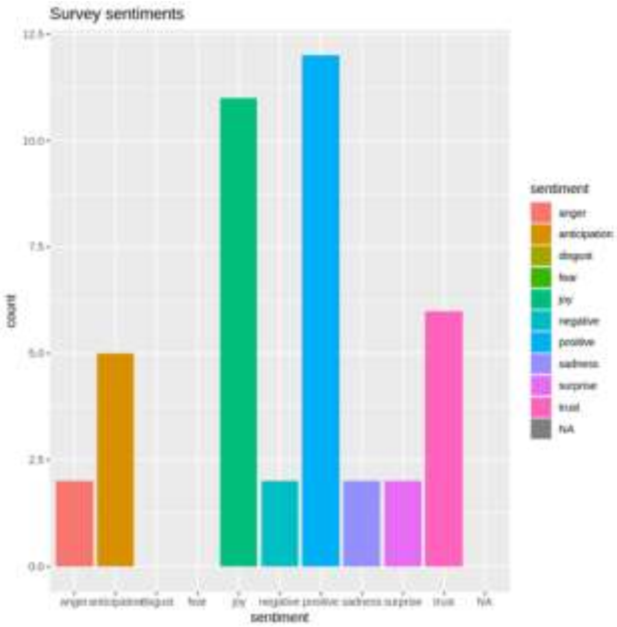

Figure 2. Emotion Analysis Results Using R

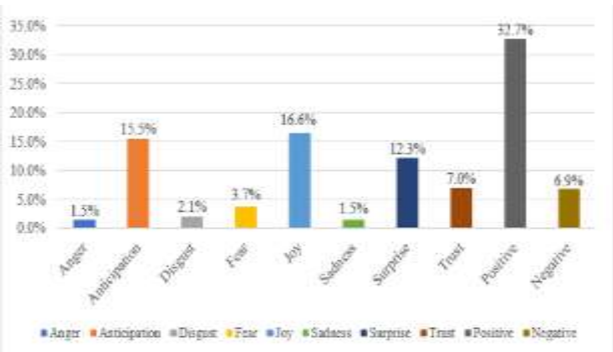

Figure 3. Emotional Score Survey Results

Based on Figure 3, positive emotions dominate, with the highest scores recorded in the Positive category at 32.7% and Joy at 16.6%. This indicates that the majority of TikTok users responded positively to the content and products of Camille Beauty Official, suggesting that the brand maintains a favorable image in the eyes of its customers. Anticipation (15.5%) and Surprise (12.3%) further suggest curiosity and interest toward new products or promotional offers. The Trust score of 7.0% reflects a degree of customer confidence in the brand, although there is room for improvement through enhanced credibility and transparency.

Conversely, negative emotions were also present, albeit in smaller proportions. The Negative category scored 6.9%, with Fear accounting for 3.7%, Disgust 2.1%, and Anger 1.5%. These findings indicate the presence of dissatisfaction related to certain aspects such as product quality, pricing, or service. Sadness registered only 1.5%, signifying that deeply negative complaints were rare.

The analysis suggests that Camille Beauty Official has substantial potential to strengthen customer loyalty by leveraging positive emotions such as Joy, Anticipation, and Surprise through creative campaigns, interactive content, and the introduction of appealing new products. However, attention must also be given to negative emotions to prevent a decline in customer satisfaction. Recommended strategies include improving customer service quality, ensuring product consistency, and providing clear communication on issues that may provoke Fear or Disgust. By reinforcing Trust and minimizing negative perceptions, Camille Beauty's brand image can become more positive and competitive in the market. Table 3 presents recommendations and corrective strategies aimed at enhancing customer satisfaction and strengthening the brand image of Camille Beauty Official.

Table 3. Proposed Recommendations And Strategic Improvements For The Research Object

	Recommendation / Strategy	Associated Emotions
1	Optimize entertaining and educational content significantly to enhance engagement [20]	Joy (17%) and Anticipation (16%) dominate among the audience, indicating high interest and curiosity toward engaging content.
2	Increase positive reviews as they enhance trust and foster loyalty among both new and existing customers [21]	Trust has a relatively high score but requires reinforcement to strengthen loyalty. Positive emotions such as Positive (33%), Joy (17%), and Anticipation (16%) are dominant.
3	Implement influencer marketing to increase	Anticipation (16%) and Surprise (12%) reflect

	engagement, followers, and viewers by reaching a broader audience on social media [22]	enthusiasm toward interactive activities.
4	Enhance customer experience and user interaction as these influence electronic word-of-mouth (eWOM), which is essential for improving brand visibility and trust [21]	Negative emotions such as Fear (3.7%) and Disgust (2.1%) should be addressed, while Trust (7%) needs to be strengthened.

IV. CONCLUSION AND IMPLICATIONS

This study investigated customer engagement on TikTok product videos of Camille Beauty Official, focusing on sentiment polarity and emotional classification derived from user-generated comments. Through systematic text preprocessing, sentiment labeling using a lexicon-based approach, and emotion analysis utilizing the NRC EmoLex lexicon, the research identified dominant sentiment patterns and associated emotional states.

The findings indicate that positive sentiments dominate (63.7%), significantly surpassing negative sentiments (36.3%), suggesting an overall favorable perception of Camille Beauty's TikTok content. Emotion classification further revealed that Positive (33%), Joy (17%), and Anticipation (16%) were the most prominent emotional categories, reflecting optimism, enthusiasm, and curiosity toward the brand's offerings. Additionally, Surprise (12%) and Trust (7%) demonstrate substantial audience engagement and brand confidence. In contrast, negative emotions such as Fear (3.7%), Disgust (2.1%), and Anger (1.5%)—though less prevalent—signal areas of potential dissatisfaction related to product quality, pricing, or service experiences.

V. LIMITATIONS AND FUTURE RESEARCH

Despite its contributions, this study has limitations that warrant consideration. First, the analysis was confined to TikTok comments on a single brand's account, which may restrict generalizability across platforms or industries. Second, the reliance on lexicon-based sentiment and emotion classification may overlook context-specific nuances such as sarcasm or mixed emotions. Future research could adopt advanced natural language processing (NLP) techniques, such as deep learning models (e.g., BERT or LSTM), to improve classification accuracy and contextual understanding. Additionally, comparative studies across multiple social media platforms or brands could provide broader insights into emotional engagement patterns. Incorporating behavioral metrics such as click-through rates, purchase intentions, and conversion data may also enhance the predictive value of sentiment-emotion analysis.

REFERENCES

1. S. R. Agrawal, "Digital Pollution and Its Impact on the Family and Social Interactions," J. Fam. Issues, vol. 42, no. 11, pp. 2648–2678, Nov. 2021, doi: 10.1177/0192513X20985558.
2. Austerberry David, "The Technology of Video and Audio Streaming Second Edition," 2005.
3. Q. Gaus, A. Jolliff, and M. A. Moreno, "A content analysis of YouTube depression personal account videos and their comments," Comput. Hum. Behav. Rep., vol. 3, Jan. 2021, doi: 10.1016/j.chbr.2020.100050.
4. G. C. Wicaksana and E. Ardyan, "Emotional Experiences Drive Customer Loyalty in Indonesia," Acad. Open, vol. 9, no. 2, July 2024, doi: 10.21070/acopen.9.2024.9144.
5. N. Nawiyah, R. C. Kaemong, M. A. Ilham, and F. Muhammad, "PENYEBAB PENGARUHNYA PERTUMBUHAN PASAR INDONESIA TERHADAP PRODUK SKIN CARE LOKAL PADA TAHUN 2022," ARMADA J. Penelit. Multidisiplin, vol. 1, no. 12, pp. 1390–1396, Dec. 2023, doi: 10.55681/armada.v1i12.1060.
6. R. Sustaningrum, "PURCHASE INTENTION GENERASI Z PADA PRODUK KOSMETIK DENGAN TEKNOLOGI AUGMENTED REALITY PADA MASA PANDEMI," J. Ilm. Ekon. Bisnis, vol. 28, no. 3, pp. 405–419, 2023, doi: 10.35760/eb.2023.v28i3.7170.
7. E. S. C. Nangoy, J. R. E. Tampi, and T. M. Tumbel, "Pemanfaatan Aplikasi Tiktok sebagai Digital Marketing Promotion pada Curabeauty Manado," Productivity, vol. 5, no. 2, pp. 859–863, June 2024, doi: 10.35797/ejp.v5i1.54704.
8. R. Fauzan Abdillah and A. N. Pramesti, "SEMINAR NASIONAL AMIKOM SURAKARTA (SEMNASA) 2024 DAMPAK RATING DAN ULASAN KONSUMEN TERHADAP KEPUTUSAN PEMBELIAN DI E-COMMERCE".
9. D. Röchert, G. K. Shahi, G. Neubaum, B. Ross, and S. Stieglitz, "The Networked Context of COVID-19 Misinformation: Informational Homogeneity on YouTube at the Beginning of the Pandemic," Online Soc. Netw. Media, vol. 26, Nov. 2021, doi: 10.1016/j.osnem.2021.100164.
10. W. Tafesse, "YouTube marketing: how marketers' video optimization practices influence video views," Internet Res., vol. 30, no. 6, pp. 1689–1707, Oct. 2020, doi: 10.1108/INTR-10-2019-0406.
11. A. Pansari and V. Kumar, "Customer engagement: the construct, antecedents, and consequences," J. Acad. Mark. Sci., vol. 45, no. 3, pp. 294–311, May 2017, doi: 10.1007/s11747-016-0485-6.
12. L. Dessart and V. Pitardi, "How stories generate

consumer engagement: An exploratory study," J. Bus. Res., vol. 104, pp. 183–195, Nov. 2019, doi: 10.1016/j.jbusres.2019.06.045.

13. S. Moldovan, Y. Steinhart, and D. R. Lehmann, "Propagators, Creativity, and Informativeness: What Helps Ads Go Viral," J. Interact. Mark., vol. 47, pp. 102–114, Aug. 2019, doi: 10.1016/j.intmar.2019.02.004.
14. A. N. Smith, E. Fischer, and C. Yongjian, "How Does Brand-related User-generated Content Differ across YouTube, Facebook, and Twitter?," J. Interact. Mark., vol. 26, no. 2, pp. 102–113, May 2012, doi: 10.1016/j.intmar.2012.01.002.
15. S. R. Agrawal and D. Mittal, "Optimizing customer engagement content strategy in retail and E-tail: Available on online product review videos," J. Retail. Consum. Serv., vol. 67, July 2022, doi: 10.1016/j.jretconser.2022.102966.
16. S. M. Mohammad and P. D. Turney, "CROWDSOURCING A WORD-EMOTION ASSOCIATION LEXICON," 2012. [Online]. Available: http://crowdsourcing.typepad.com/cs/2006/06
17. S. Chatterjee, "Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents," Decis. Support Syst., vol. 119, pp. 14–22, Apr. 2019, doi: 10.1016/j.dss.2019.02.008.
18. Y. Dang, Y. Zhang, and H. Chen, "A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews," 2010. [Online]. Available: www.computer.org/intelligent
19. M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," 2011.
20. S. Febriyanti and A. Budiman, "Pendampingan Peningkatan Kreativitas Konten Digital Sebagai Upaya Meningkatkan Engagement Media Sosial Pada 86 Production".
21. Aris Kurniawan, Lili Marlinah, Yosie Noverha, and Vina Islami, "Pengaruh Digital Marketing, Social Media Engagement, dan Customer Trust Terhadap Loyalitas Pelanggan E-Commerce," El-Mal J. Kaji. Ekon. Bisnis Islam, vol. 5, no. 10, Oct. 2024, doi: 10.47467/elmal.v5i10.5268.
22. F. Wardah, "Analisis Pengaruh Influencerterhadap Minat Beli Konsumen pada Perusahaan JavaMifi," vol. 02, no. 03, 2023.