KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 4, No 6, June 2024, Hal 2937-2946
DOI 10.30865/klik.v4i6.1946
https://djournals.com/klik

Sentiment and Toxicity Score Evaluation of DJI Avata Product Reviews Using Cross-Industry Standard Process for Data Mining

Yerik Afrianto Singgalen

Faculty of Business Administration and Communication, Tourism Department, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: yerik.afrianto@atmajaya.ac.id
Email Corresponding Author: yerik.afrianto@atmajaya.ac.id

Abstract−This research employs the CRISP-DM framework to analyze consumer sentiment and preferences regarding DJI Avata drone products, aiming to provide data-driven strategic recommendations for marketing and product development. By systematically exploring business objectives, preparing and cleaning data, and modeling sentiment, the study reveals high consumer engagement and predominantly positive sentiment (51.91% positive, 31.16% neutral, 16.93% negative) towards the DJI Avata. The Support Vector Machine (SVM) algorithm demonstrated superior performance in sentiment classification, achieving an accuracy of 74.69%, with an AUC of 0.839, precision of 77.57%, recall of 69.68%, and F-measure of 73.23%. A comparative analysis between the VADER and TextBlob models, showing a moderate agreement (Cohen's kappa statistic = 0.413) on 64.84% of the posts, highlighted the value of using multiple sentiment analysis tools. Furthermore, toxicity scores calculated via the Perspective API identified critical areas for improvement in user engagement. Subsequently, the toxicity results reveal the following scores: Toxicity with an average of 0.09461 and a peak of 0.90451, Severe Toxicity with an average of 0.00817 and a peak of 0.45895, Identity Attack with an average of 0.01139 and a peak of 0.58743, Insult with an average of 0.04543 and a peak of 0.70658, Profanity with an average of 0.06133 and a peak of 0.89080, and Threat with an average of 0.02063 and a peak of 0.69437. These detailed metrics provide a comprehensive understanding of the dataset's different dimensions and intensities of negative sentiments. The significant variation between average and peak values indicates the presence of highly negative interactions, which necessitates targeted intervention. Consequently, these findings inform the development of specific strategies to mitigate toxicity and enhance the overall user experience in digital communities. These insights informed strategic recommendations to enhance digital marketing efforts and product features, underscoring the CRISP-DM framework's efficacy in guiding comprehensive consumer sentiment analysis and fostering informed decision-making in the aerial photography and videography market.

Keywords: CRISP-DM Framework; Consumer Sentiment Analysis; DJI Avata; Digital Marketing; Sentiment Classification

1. INTRODUCTION

Aerial photography and videography products, such as drones, have become indispensable tools for content creators in the digital marketing era. This surge in drone utilization has sparked considerable consumer commentary regarding the various brands available [1]. The widespread adoption of drones is driven by the ability to capture high-quality, immersive visuals that significantly enhance marketing campaigns [2]–[4]. Despite the initial investment, the return on visual content engagement justifies the expense, highlighting the critical role of drones in contemporary content creation [5]. Consequently, the increasing consumer discourse around drone brands underscores the integral role in digital marketing strategies.

A significant challenge in developing product and service marketing studies lies in contextualizing products based on the targeted market segments, exemplified by content creators' use of drones for aerial photography and videography [6]–[9]. Addressing this challenge involves understanding different consumer segments' distinct preferences and requirements, directly influencing marketing strategies [10]–[15]. Additionally, the dynamic nature of digital content creation necessitates continuous adaptation and innovation in product offerings [16]–[20]. Effectively contextualizing products ensures the relevance and appeal to specific market segments, ultimately driving success in competitive markets.

Based on the market segment for drone products, influencers conducting product reviews through audiovisual content stimulate purchase intentions via various online channels. These influencers leverage credibility and reach to showcase product features and benefits, engaging potential buyers [21], [22]. Moreover, such content's visual and auditory appeal provides a comprehensive understanding of the product's capabilities, enhancing consumer confidence [23]–[25]. Consequently, influencer reviews significantly impact consumer purchasing behavior, highlighting the importance of strategic influencer partnerships in digital marketing efforts.

The urgency of this research is underscored by the rapid proliferation of digital content and the consequential rise in user-generated reviews and feedback. Understanding consumer sentiment and managing toxicity in online interactions are critical for maintaining a positive digital environment. This study's comprehensive analysis of sentiment and toxicity provides actionable insights that can significantly enhance community management and product development. Promptly addressing negative sentiments and toxicity is essential to foster a healthier, more engaging user experience. Consequently, the findings of this research offer timely and practical solutions to pressing challenges in digital marketing and online community management.

This study uses the CRISP-DM methodology to analyze sentiment and calculate the toxicity score of the DJI Avata aerial photography and videography product. By systematically applying this data mining process, sentiment analysis will reveal consumer perceptions, while the toxicity score will quantify the presence of harmful language in user reviews. Employing CRISP-DM ensures a structured and comprehensive analysis, facilitating accurate insights into consumer sentiment [26]. Ultimately, this research provides valuable data for understanding market reception and guiding product improvement strategies.

The theoretical and practical contributions of this research are significant and multifaceted. Theoretically, it advances knowledge by offering new insights into consumer behavior and sentiment analysis within the context of aerial photography and videography products [27], [28]. The findings provide actionable intelligence for marketers and product developers to refine strategies and enhance product offerings based on data-driven insights [29], [30]. This dual contribution enriches the academic literature and directly benefits industry stakeholders by informing effective decision-making processes. Consequently, this research bridges the gap between theory and practice, fostering innovation and improvement in the field.

The limitation of this research lies in the specificity of the methodology and product description through audiovisual channels and content. This focus may restrict the generalizability of the findings to other products or contexts, as the analysis is heavily dependent on the unique characteristics of the audiovisual medium. Additionally, the reliance on audiovisual content might not fully capture the breadth of consumer sentiment expressed through other forms of media. Therefore, while the insights gained are valuable, they should be interpreted with caution and considered within the specific scope of the study. This limitation highlights the need for further research incorporating diverse methodologies and broader contexts.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis

Gap analysis is essential for understanding consumer perceptions of aerial photography and videography products marketed through YouTube. This method identifies discrepancies between current knowledge and potential insights, highlighting areas where further investigation is necessary. Evaluating consumer feedback on YouTube offers a nuanced understanding of market reception and user satisfaction. Thus, conducting a thorough gap analysis illuminates existing research shortcomings and guides future studies in addressing these gaps, ultimately enhancing the effectiveness of marketing strategies on digital platforms.

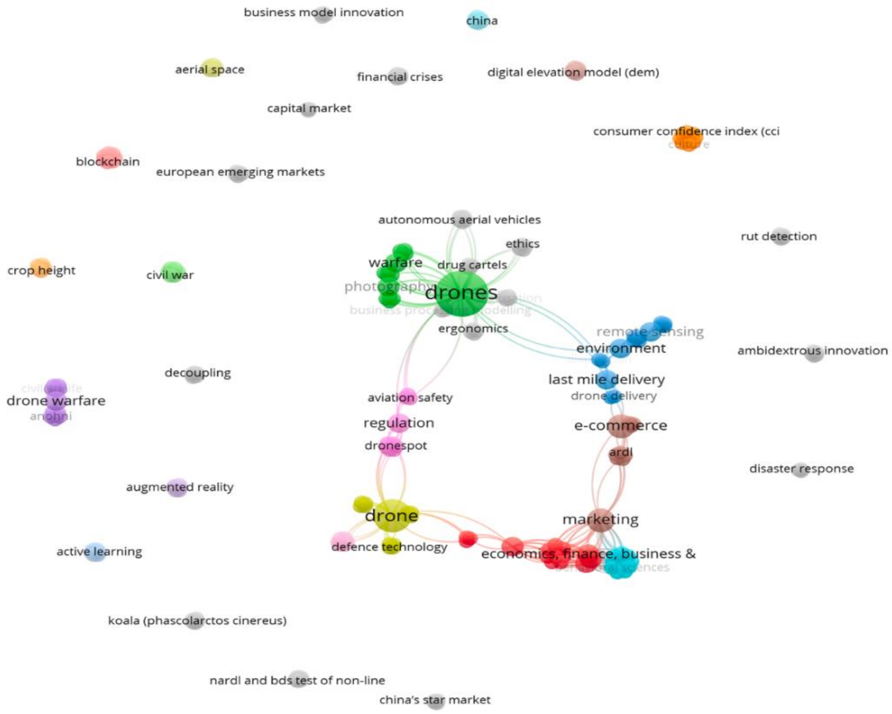

Figure 1. Gap Analysis using Vosviewer

Figure 1 shows the network, overlay, and density visualization. Based on the results of the gap identification, it is evident that drone products used in aerial photography and videography have not been extensively studied from the perspectives of consumer perception, consumer behavior, and purchasing behavior. Current research primarily employs a generalized theoretical framework, focusing on product patterns and trends rather than specific consumer insights [21], [22], [31]. This oversight suggests a need for more targeted studies that delve into the nuanced consumer interactions with drone technology. Addressing these gaps will provide a more comprehensive understanding of market dynamics and inform more effective marketing strategies tailored to consumer needs.

2.2 Cross-Industry Standard Process for Data-Mining (CRISP-DM)

The CRISP-DM framework is employed to understand business context, data, models, evaluation, and marketing optimization recommendations during deployment. This structured approach facilitates a comprehensive analysis of business objectives and relevant data, ensuring that models are accurately developed and assessed. Furthermore, the framework's iterative process allows for continuous refinement, resulting in actionable insights and effective marketing strategies. Consequently, the CRISP-DM framework's systematic methodology significantly enhances the precision and efficacy of marketing optimizations, ensuring alignment with business goals and market demands.

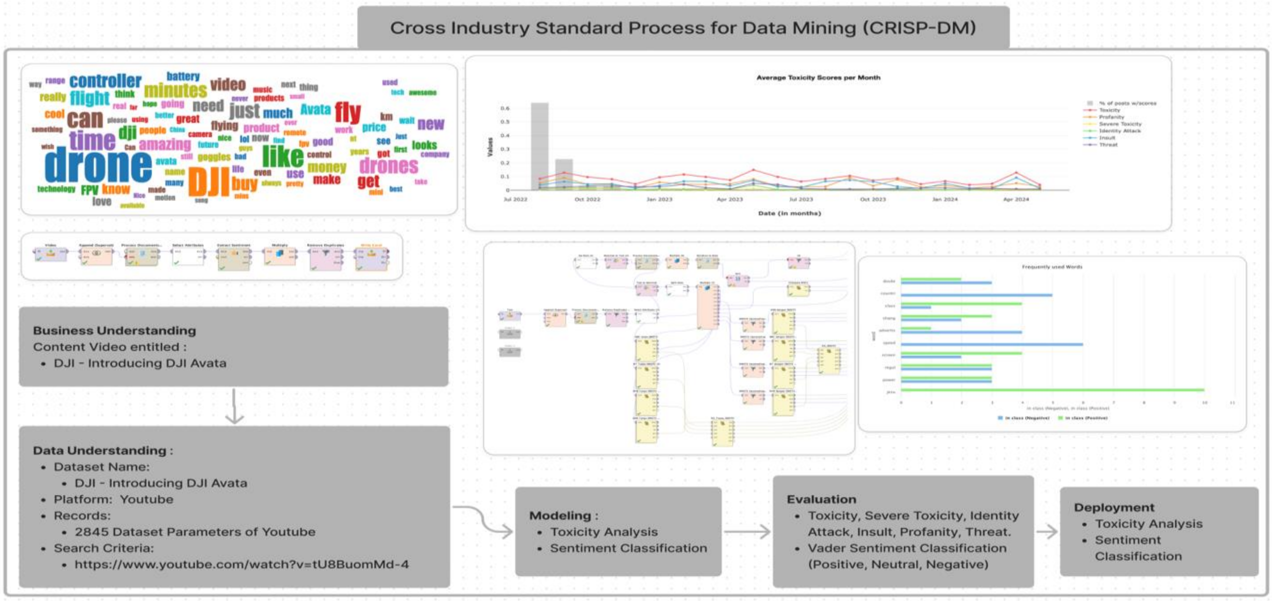

Figure 2. Implementation of CRISP-DM Framework

Figure 2 shows the implementation of the CRISP-DM framework. Implementing the CRISP-DM framework enables a comprehensive analysis of consumer perceptions and preferences regarding drone products, particularly the DJI Avata. This methodical approach systematically processes business understanding, data collection, and model development to uncover valuable insights. By evaluating these insights, the framework facilitates the generation of well-informed recommendations for product development and digital marketing optimization. Therefore, applying CRISP-DM enhances the accuracy of consumer analysis and drives strategic decisions that align with market demands, ultimately improving both product offerings and marketing effectiveness.

2.2.1 Business Understanding

At the business understanding stage, the promotional video for the DJI Avata, titled "DJI - Introducing DJI Avata," was published through the official DJI account and has garnered 3,972,921 views since its premiere on August 25, 2022. This high view count reflects significant consumer interest and engagement with the product. Such metrics indicate the effectiveness of the promotional strategy in capturing the target audience's attention. Consequently, analyzing this engagement provides valuable insights into consumer behavior and preferences, guiding further marketing and product development strategies.

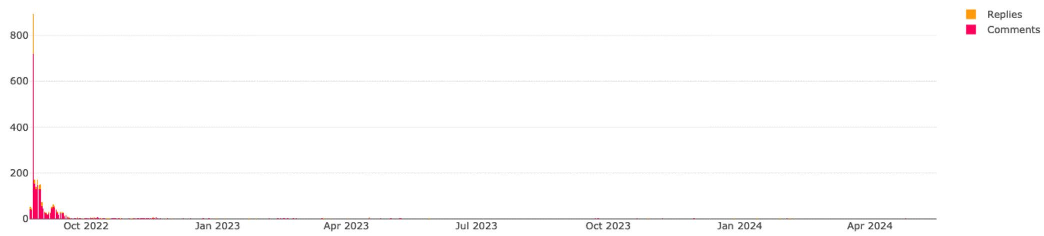

Figure 3. Video and Post-Per-Day Statistic (Communalytic)

Figure 2 shows the post-per-day statistic. Based on statistical data, the dates and number of comments on the DJI Avata promotional video are as follows: August 25, 2022, with 720 comments; August 26, 2022, with 152 comments; August 28, 2022, with 146 comments; August 29, 2022, with 130 comments; August 30, 2022, with 130 comments; August 27, 2022, with 129 comments; August 31, 2022, with 56 comments; September 8, 2022, with 52 comments; September 9, 2022, with 48 comments; and September 7, 2022, with 46 comments. This distribution reveals a significant initial spike in engagement followed by a gradual decline, suggesting that the video's release generated immediate interest, which tapered over time. Such patterns highlight the importance of timing and initial promotional efforts in maximizing consumer interaction. Consequently, this data informs future marketing strategies to sustain engagement beyond the initial release period.

Based on the number of responses to the DJI Avata drone product video with ID tU8BuomMd-4, it is evident that there has been a significant increase in interest in drones used in aerial photography and videography. This heightened engagement is reflected in the substantial number of comments and interactions, indicating a growing consumer fascination with advanced drone technology. Such trends suggest that the market for these products is expanding rapidly, driven by consumer demand for high-quality visual content. Consequently, this data underscores the importance of continued innovation and strategic marketing to capitalize on this increasing interest.

2.2.2 Data Understanding

At the data understanding stage, it is crucial to meticulously undertake the data cleaning and review extraction processes, followed by modeling to achieve high accuracy in review classification. Effective data cleaning ensures that the dataset is free from inconsistencies and irrelevant information, enhancing the quality of the extracted reviews. Subsequent modeling using advanced algorithms is pivotal in accurately classifying the reviews into meaningful categories. This systematic approach not only improves the reliability of the data analysis but also provides valuable insights for informed decision-making. Hence, thorough data preparation and precise modeling are fundamental to achieving optimal classification accuracy.

Furthermore, it is essential to understand the characteristics of the data to be analyzed from the top ten posters that provide information about the channels or accounts that contribute most significantly to reviews and responses to other users' reviews. Identifying these key contributors allows a more targeted analysis of influential opinions and engagement patterns. This focus on high-impact users aids in recognizing the dynamics of user interaction and the spread of information within the community. Consequently, analyzing the data from these top contributors provides deeper insights into the overall sentiment and engagement trends, thereby enhancing the effectiveness of subsequent analytical strategies.

Figure 4. Top-Ten-Poster (Communalytic)

Figure 4 shows the top ten posters of the dataset. Based on the data from the top-ten posters, it is evident that @Gamingbot0-text contributed 53 posts, followed by @TRICKYBYRD with eight posts, @silbay and @DUVALTV each with seven posts, @phaenosmusic with six posts, and several users including @youtubeisproCCP, @Ezzywheels, @SuperKREPSINIS, and @STEELCITYDRONEPILOT, each contributing five posts. Given the substantial volume of contributions, this distribution highlights the significant influence of @Gamingbot0-text in shaping the discourse around the product. Understanding the activity levels of these critical accounts provides insights into the sources of dominant perspectives and engagement within the community. Consequently, this information is crucial for identifying influential voices and tailoring marketing strategies accordingly.

Subsequently, the data cleaning process was conducted using RapidMiner, employing operators such as tokenize, transform cases, stopwords, stem, and other supportive pre-processing tools. The tokenize operator breaks down the text into manageable units while transforming cases standardized text into a consistent format. Removing stopwords and stemming further refines the data, eliminating noise and reducing words to the base forms. This meticulous pre-processing ensures the dataset is clean and ready for accurate analysis. Therefore, leveraging these RapidMiner operators significantly enhances the quality and reliability of the data, facilitating more precise and insightful results.

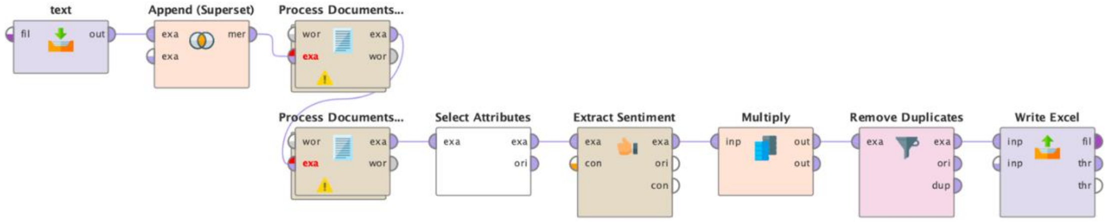

Figure 5. Data Cleaning Process (Rapidminer)

Figure 5 shows the data cleaning process in Rapidminer. After the data has been cleaned, the extract sentiment operator is utilized through the VADER model to label the data based on positive and negative classes. This sentiment analysis tool efficiently categorizes textual data, distinguishing between favorable and unfavorable sentiments. Applying the VADER model systematically labels the data, enhancing the accuracy of subsequent analysis. Consequently, this process facilitates a deeper understanding of consumer opinions and trends, contributing to more informed decision-making and strategy development.

Labeled data then proceed to the modeling process to determine the best-performing model for classification. This step involves training various models to evaluate the accuracy, precision, and recall in categorizing the data. The most effective model for the given dataset was identified by comparing these metrics, ensuring robust and reliable classification. Consequently, selecting the optimal model enhances the analysis's overall predictive performance and reliability, supporting more accurate and actionable insights.

2.2.3 Modeling

At the modeling stage, toxicity scores were calculated to identify the average and highest values for categories such as Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, and Threat. By processing data from 2,447 posts (out of 2,845) using the Perspective API, the analysis aimed to gauge the intensity and prevalence of negative sentiment within the dataset. These toxicity metrics provide a nuanced understanding of the potential harmful interactions in user comments. Consequently, the findings highlight areas where community management and moderation efforts improved, ensuring a safer and more positive online environment.

A limitation of the Perspective API model is its language support scope. The Perspective models currently support Arabic, Chinese, Czech, Dutch, English, French, German, Hindi, Hinglish, Indonesian, Italian, Japanese, Korean, Polish, Portuguese, Russian, Spanish, and Swedish. This restricted language range may impact the model's applicability and effectiveness in analyzing multilingual datasets. Consequently, the insights derived from non-supported languages may be less accurate or entirely unavailable, potentially limiting the comprehensiveness of sentiment analysis in a global context. Addressing this limitation requires additional language support to enhance the model's versatility and utility.

In addition, algorithms such as k-NN, SVM, DT, and NBC are tested for classification. Each algorithm's performance is evaluated based on accuracy, precision, and recall metrics. The most effective algorithm is identified through rigorous testing, which will be discussed in detail in this article. This thorough evaluation ensures that the chosen algorithm provides the highest classification accuracy, supporting robust and reliable data analysis. Consequently, focusing on the best-performing algorithm enhances the overall validity and applicability of the research findings.

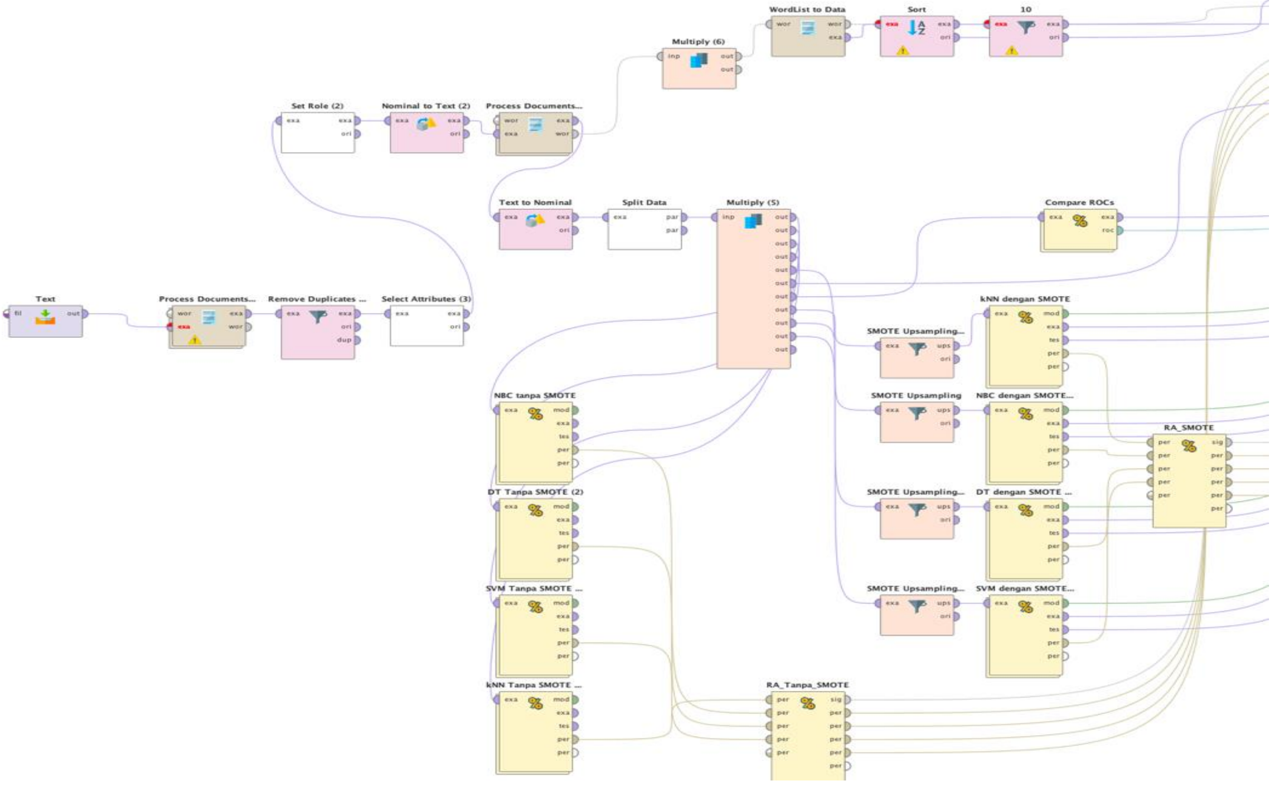

Figure 6. Implementation of SVM and DT Models in Rapidminer

Figure 6 shows the modeling process using Rapidminer. The modeling results will be evaluated based on accuracy, precision, recall, AUC, and F-measure. These metrics comprehensively assess the model's performance, capturing different aspects of its classification ability. Accuracy measures the overall correctness of the model, while precision and recall evaluate its performance in identifying positive instances. The AUC metric assesses the model's ability to distinguish between classes, and the F-measure provides a balanced evaluation of precision and recall. Therefore, these metrics ensure a thorough and nuanced evaluation of the model's effectiveness in classification tasks.

The algorithm with the best performance will be recommended as the most relevant model for processing review data of DJI Avata product content. This recommendation is based on thoroughly evaluating various metrics, ensuring the selected algorithm demonstrates superior accuracy, precision, recall, AUC, and F-measure. The model's robustness and reliability in classifying consumer reviews are crucial for gaining actionable insights. Consequently, implementing the top-performing algorithm will enhance the effectiveness of data analysis and support strategic decision-making in digital marketing efforts for DJI Avata.

2.2.4 Evaluation

At the evaluation stage, the results of the toxicity score were analyzed comprehensively to generate precise recommendations. This thorough analysis examined the average and highest values of various toxicity categories, such as Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, and Threat. Understanding these metrics allowed for a detailed assessment of the harmful sentiment levels within the dataset. Consequently, this informed the development of targeted strategies for community management and content moderation. Ultimately, such a meticulous evaluation ensures that the recommendations are accurate and actionable, enhancing the overall user experience and promoting a healthier online environment.

Subsequently, the toxicity results were evaluated based on the classification of negative, neutral, and positive sentiments to generate precise recommendations. This comprehensive evaluation involved categorizing the sentiment scores and analyzing the distribution and intensity of toxicity across these classifications. More accurate insights into user feedback and behavior were obtained by understanding the prevalence of negative, neutral, and optimistic sentiments. Consequently, this analysis informed the development of targeted strategies for community engagement and content moderation. Ultimately, this systematic approach ensures that the recommendations are well-informed and effective, enhancing the overall quality and safety of the online environment.

2.2.5 Deployment

At the deployment stage, it is recommended that a marketing strategy for drone products be implemented aimed at aerial photography and videography for content creators. This strategy should leverage insights gained from data analysis to target specific consumer segments effectively. Utilizing platforms popular among content creators like YouTube and Instagram maximizes engagement and reach. Additionally, collaborations with influencers with a substantial following in the photography and videography community further amplify the product's visibility. Consequently, these targeted marketing efforts will likely enhance brand awareness and drive sales within this niche market.

3. RESULT AND DISCUSSION

The discussion in this research indicates that the toxicity score, derived using the Perspective API model, provides a comprehensive classification of review data based on categories such as Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, and Threat. This detailed categorization allows for a nuanced understanding of the nature and intensity of negative sentiments within user reviews. The model's ability to distinguish between various toxicity levels and types enhances sentiment analysis's accuracy. Consequently, these insights are instrumental in developing strategies to address and mitigate negative feedback, improving overall user engagement and satisfaction.

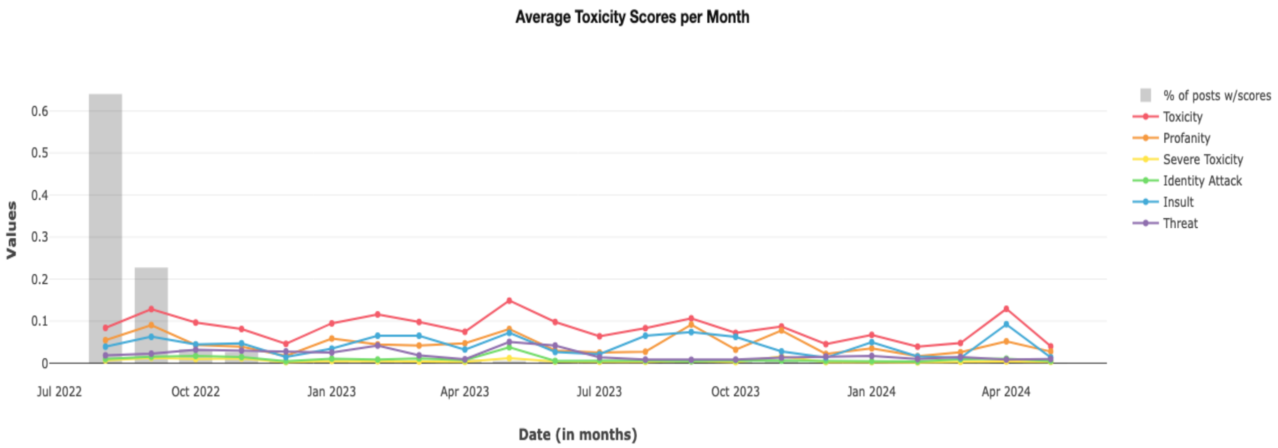

Figure 7. Average Toxicity Scores per Month

Figure 7 shows the average Toxicity Scores per Month. Based on the average toxicity score per month, it is observed that the mean and highest values of toxicity are as follows: Toxicity with an average of 0.09461 and a highest value of 0.90451, Severe Toxicity with an average of 0.00817 and a highest value of 0.45895, Identity Attack with an average of 0.01139 and a highest value of 0.58743, Insult with an average of 0.04543 and a highest value of 0.70658, Profanity with an average of 0.06133 and the highest value of 0.89080, and Threat with an average of 0.02063 and the highest value of 0.69437. These metrics provide a detailed view of user reviews' varying degrees of negative sentiment. The significant disparity between the average and peak values underscores the potential for extreme negativity in certain instances. Consequently, these insights are crucial for developing targeted strategies to address specific types of toxicity, ultimately enhancing the user experience and fostering a more positive online community.

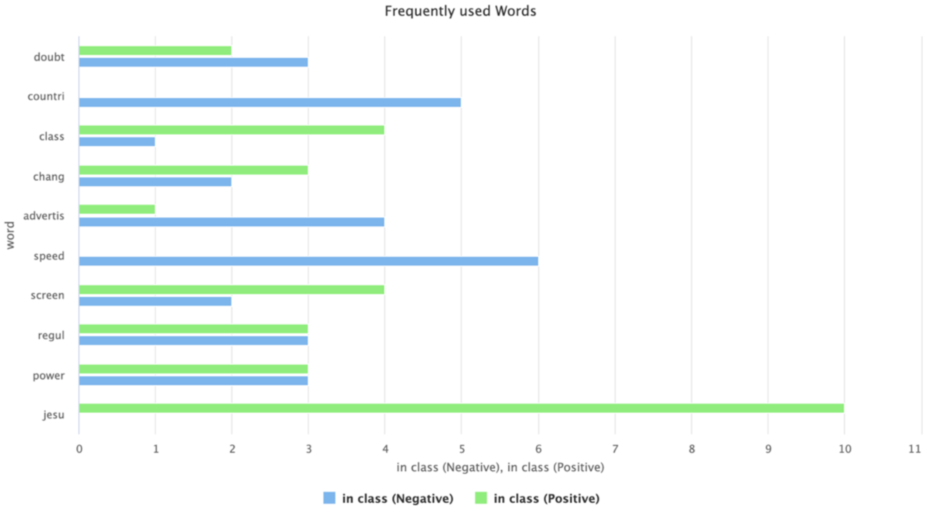

Figure 8. Frequently used Words (Rapidminer)

Figure 8 shows the frequently used words in the dataset. The frequently used words indicate the occurrence in documents and totals as follows: "Jesu" appears three times in documents with a total of 10 occurrences, "power" 3 times in documents with six occurrences, "regular" 4 times in documents with six occurrences, "screen" 4 times in documents with six occurrences, "speed" 3 times in documents with six occurrences, "advertise" 4 times in documents with five occurrences, "change" 3 times in documents with five occurrences, "class" 3 times in documents with five occurrences, "country" 4 times in documents with five occurrences, and "doubt" 4 times in documents with five occurrences. This frequency analysis highlights the key themes and concepts prevalent in the documents. The repetition of these words underscores the significance in the context of the text, suggesting they are central to the discussion. Consequently, understanding the frequency and distribution of these terms provides valuable insights into the dataset's main topics and focus areas.

There is a discrepancy in the calculation and identification of frequently used words between RapidMiner and Communalytic. RapidMiner's analysis reveals a distinct set of keywords, emphasizing terms that might reflect specific trends or patterns unique to its algorithms. Conversely, Communalytic's results highlight frequently used words, potentially due to its varied methodological approach in text mining and data analysis. This divergence in findings suggests that the choice of tool significantly influences the outcomes of text analysis. Consequently, it is crucial to consider these differences when interpreting the results, as they may impact the overall insights and conclusions drawn from the data.

Figure 9. Frequently Used Words (Communalytic)

Figure 9 shows the frequently used words based on Communalytic. The identification of frequently used words on Communalytic reveals the following terms and frequencies: "drone" appears 141 times, "DJI" 114 times, "like" 81 times, "fly" 76 times, "time" 73 times, "can" 70 times, "drones" 57 times, "just" 54 times, "minutes" 52 times, "flight" 49 times, "controller" 48 times, "buy" 48 times, "DJI" 46 times, "get" 44 times, "new" 42 times, "video" 41 times, "amazing" 40 times, "need" 39 times, and "money" 36 times. These frequencies highlight the prominence of specific themes and concepts within the dataset, particularly those related to drone technology and consumer experiences. The recurrent appearance of terms such as "fly," "flight," and "controller" underscores the technical aspects of drone usage. At the same time, words like "buy," "get," and "money" reflect consumer behavior and purchasing considerations. Consequently, these insights provide a nuanced understanding of users' central topics and concerns about DJI drones.

Subsequently, the sentiment analysis modeling results on Communalytic, which utilized the VADER model to process 2,240 out of 2,845 posts, reveal the following distribution: out of 2,150 posts, 364 (16.93%) exhibited negative sentiment, 670 (31.16%) showed neutral sentiment, and 1,116 (51.91%) expressed positive sentiment. These findings indicate a predominance of positive sentiment within the dataset, suggesting a generally favorable reception of the topics discussed. The significant proportion of neutral sentiments highlights a considerable amount of impartial or balanced discourse. Consequently, the analysis provides valuable insights into the overall sentiment landscape, informing more nuanced interpretations and strategic decisions based on the data.

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	2150	364 (16.93%)	670 (31.16%)	1116 (51.91%)
TextBlob (English/EN)	2150	223 (10.37%)	805 (37.44%)	1122 (52.19%)
TextBlob (French/FR)	26	2 (7.69%)	21 (80.77%)	3 (11.54%)
TextBlob (German/DE)	30	1 (3.33%)	27 (90.00%)	2 (6.67%)

Figure 10. Implementation of Vader Model in Sentiment Analysis

Figure 10 shows the per. Based on the comparison between the VADER and TextBlob models, it is evident that both models agree on the categorization of 1,348 (64.84%) out of 2,079 English language posts. This level of agreement is considered moderate, as indicated by a Cohen's kappa statistic of 0.413. Specifically, both models concur on the following: 121 (8.98%) posts with negative sentiments (polarity scores <= -0.05), 445 (33.01%) posts with neutral sentiments (polarity scores between -0.05 and 0.05), and 782 (58.01%) posts with positive sentiments (polarity scores >= 0.05). These agreements are highlighted with a green background in the confusion matrix below, illustrating a significant alignment in sentiment categorization between the two models. Consequently, the moderate agreement level underscores the importance of using multiple models to analyze sentiment comprehensively.

Subsequently, the modeling results indicate that the Support Vector Machine (SVM) algorithm best classifies sentiment data in RapidMiner. The PerformanceVector shows an accuracy of 74.69% +/- 8.13% (micro average: 74.70%). The ConfusionMatrix details the true negatives as 200 and the true positives as 175, with false negatives and positives being 51 and 76, respectively. The optimistic AUC is 0.844 +/- 0.065 (micro average: 0.844) for the positive class, while the regular and pessimistic AUCs are 0.839 +/- 0.063 (micro average: 0.839) and 0.833 +/- 0.061 (micro average: 0.833), respectively. The precision is 77.57% +/- 9.29% (micro average: 77.43%), and the recall is 69.68% +/- 9.89% (micro average: 69.72%). The f_measure is 73.23% +/- 8.79% (micro average: 73.38%). These metrics collectively demonstrate the robustness and reliability of the SVM algorithm in sentiment classification, making it a highly effective model for this task.

The recommendations based on the results of this research are multifaceted and aim to enhance product development and marketing strategies for DJI Avata drone products. First, leveraging the Support Vector Machine (SVM) model for sentiment classification has proven effective and should be incorporated into ongoing data analysis efforts to monitor consumer sentiment accurately. Second, addressing the identified toxicity issues through improved community management and moderation fosters a more positive user environment. Additionally, expanding language support in sentiment analysis tools, such as the Perspective API, will enable a more comprehensive analysis of global consumer feedback. Consequently, these steps will ensure a more robust understanding of consumer needs and preferences, driving better-informed decision-making and strategic planning.

4. CONCLUSION

This research utilized the CRISP-DM framework to analyze consumer sentiment and preferences regarding DJI Avata drone products, leading to significant insights and strategic recommendations. The structured approach encompassed business understanding, data preparation, and modeling, revealing high consumer interest and predominantly positive sentiment towards the product (51.91% positive, 31.16% neutral, 16.93% unfavorable). The Support Vector Machine

(SVM) algorithm emerged as the most effective for sentiment classification, with an accuracy of 74.69% and robust performance metrics, including an AUC of 0.839, precision of 77.57%, recall of 69.68%, and F-measure of 73.23%. Additionally, the comparison between VADER and TextBlob models, which agreed on 64.84% of the 2,079 English language posts, underscored the importance of using multiple tools for comprehensive analysis. The toxicity score analysis using the Perspective API revealed average and highest values for various toxicity categories, such as Toxicity (0.09461, 0.90451) and Severe Toxicity (0.00817, 0.45895), indicating areas for potential improvement in community engagement. Consequently, the findings informed strategic marketing recommendations aimed at optimizing digital marketing efforts and enhancing product features to better cater to consumer needs, thereby demonstrating the framework's efficacy in guiding data-driven decision-making in the context of aerial photography and videography products.

ACKNOWLEDGEMENTS

I sincerely thank the Tourism Department, the Faculty of Business Administration and Communication, LPPM, the Center of Digital Transformation and Tourism Development, and the Atma Jaya Catholic University of Indonesia.

REFERENCES

[1] A. Calcara, A. Gilli, M. Gilli, and I. Zaccagnini, "Will the Drone Always Get Through? Offensive Myths and Defensive Realities," *Secur. Stud.*, vol. 31, no. 5, pp. 791–825, 2022, doi: 10.1080/09636412.2022.2153734.

[2] G. Krame, V. Vivoda, and A. Davies, "Narco drones: tracing the evolution of cartel aerial tactics in Mexico's low-intensity conflicts," *Small Wars Insur.*, vol. 34, no. 6, pp. 1095–1129, 2023, doi: 10.1080/09592318.2023.2226382.

[3] A. Zenz, "Safety first: analysing the problematisation of drones," *Griffith Law Rev.*, vol. 32, no. 3, pp. 310–334, 2023, doi: 10.1080/10383441.2024.2303937.

[4] D. Kunertova, "Drones have boots: Learning from Russia's war in Ukraine," *Contemp. Secur. Policy*, vol. 44, no. 4, pp. 576–591, 2023, doi: 10.1080/13523260.2023.2262792.

[5] C. S. Ritter, "Gazing from the air: tourist encounters in the age of travel drones," *Tour. Geogr.*, vol. 0, no. 0, pp. 1–17, 2023, doi: 10.1080/14616688.2023.2264823.

[6] A. Tkaczynski and S. Rundle-Thiele, "Koala conservation in South East Queensland: a shared responsibility," *Australas. J. Environ. Manag.*, vol. 30, no. 1, pp. 48–67, 2023, doi: 10.1080/14486563.2023.2173320.

[7] S. C. Teo, T. W. Liew, and H. Y. Lim, "Factors influencing consumers' continuance purchase intention of local food via online food delivery services: the moderating role of gender," *Cogent Bus. Manag.*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311975.2024.2316919.

[8] H. N. Chuong, V. T. P. Uyen, N. D. P. Ngan, N. T. B. Tram, L. N. B. Tran, and N. T. T. Ha, "Exploring a new service prospect: customer' intention determinants in light of utaut theory," *Cogent Bus. Manag.*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311975.2023.2291856.

[9] J. M. de Luis-Ruiz, J. Sedano-Cibrián, R. Pérez-Álvarez, R. Pereda-García, and B. Malagón-Picón, "Metric contrast of thermal 3D models of large industrial facilities obtained by means of low-cost infrared sensors in UAV platforms," *Int. J. Remote Sens.*, vol. 43, no. 2, pp. 457–483, 2022, doi: 10.1080/01431161.2021.2003903.

[10] J. Sun, S. Wang, and F. Yuan, "The relationship between intrapreneurial capabilities and development in high-tech SMEs in China," *Asian J. Technol. Innov.*, vol. 32, no. 1, pp. 160–181, 2024, doi: 10.1080/19761597.2023.2177878.

[11] K. H. Zhang, "U.S.-China Economic Links and Technological Decoupling," *Chinese Econ.*, vol. 56, no. 5, pp. 353–365, 2023, doi: 10.1080/10971475.2023.2173399.

[12] S. Kojima *et al.*, "Heterogeneous robots coordination for industrial plant inspection and evaluation at World Robot Summit 2020," *Adv. Robot.*, vol. 36, no. 21, pp. 1102–1119, 2022, doi: 10.1080/01691864.2022.2111230.

[13] S. Khodjaev, L. Kuhn, I. Bobojonov, and T. Glauben, "Combining multiple UAV-Based indicators for wheat yield estimation, a case study from Germany," *Eur. J. Remote Sens.*, vol. 57, no. 1, 2024, doi: 10.1080/22797254.2023.2294121.

[14] D. R. Burchfield, S. L. Petersen, S. G. Kitchen, and R. R. Jensen, "sUAS-Based Remote Sensing in Mountainous Areas: Benefits, Challenges, and Best Practices," *Pap. Appl. Geogr.*, vol. 6, no. 1, pp. 72–83, 2020, doi: 10.1080/23754931.2020.1716385.

[15] E. Masson-MacLean, J. O'Driscoll, C. McIver, and G. Noble, "Digitally Recording Excavations on a Budget: A (Low-Cost) DIY Approach from Scotland," *J. F. Archaeol.*, vol. 46, no. 8, pp. 595–613, 2021, doi: 10.1080/00934690.2021.1970444.

[16] J. W. Park and D. J. Yeom, "Method for establishing ground control points to realize UAV-based precision digital maps of earthwork sites," *J. Asian Archit. Build. Eng.*, vol. 21, no. 1, pp. 110–119, 2022, doi: 10.1080/13467581.2020.1869023.

[17] M. I. Tabash, U. A. Sheikh, and M. Asad, "Market miracles: Resilience of Karachi stock exchange index against terrorism in Pakistan," *Cogent Econ. Financ.*, vol. 8, no. 1, 2020, doi: 10.1080/23322039.2020.1821998.

[18] T. Farooq, S. Lucas, and S. Wolff, "Predators and Peace: Explaining the Failure of the Pakistani Conflict Settlement Process in 2013-4," *Civ. Wars*, vol. 22, no. 1, pp. 26–63, 2020, doi: 10.1080/13698249.2020.1704603.

[19] M. Heath, A. Imran, and D. St-Onge, "See as a Bee: UV Sensor for Aerial Strawberry Crop Monitoring," *Can. J. Remote Sens.*, vol. 50, no. 1, p., 2024, doi: 10.1080/07038992.2024.2332179.

[20] M. Mutschler, M. Bales, and E. Meininghaus, "The impact of precision strike technology on the warfare of non-state armed groups: case studies on Daesh and the Houthis," *Small Wars Insur.*, vol. 00, no. 00, pp. 1–28, 2024, doi: 10.1080/09592318.2024.2319216.

[21] C. Chen, S. Leon, and P. Ractham, "Will customers adopt last-mile drone delivery services? An analysis of drone delivery in the emerging market economy," *Cogent Bus. Manag.*, vol. 9, no. 1, 2022, doi: 10.1080/23311975.2022.2074340.

[22] P. Abichandani *et al.*, "Competition-based active learning instruction for drone education," *Interact. Learn. Environ.*, vol. 0, no. 0, pp. 1–19, 2022, doi: 10.1080/10494820.2022.2128821.

[23] L. A. O'Hagan and E. Serafinelli, "Transhistoricizing the Drone: A Comparative Visual Social Semiotic Analysis of Pigeon and

Domestic Drone Photography," *Photogr. Cult.*, vol. 15, no. 4, pp. 327–351, 2022, doi: 10.1080/17514517.2022.2116899.

[24] A. M. El-Adle, A. Ghoniem, and M. Haouari, "The cost of carrier consistency: Last-mile delivery by vehicle and drone for subscription-based orders," *J. Oper. Res. Soc.*, vol. 75, no. 5, pp. 821–840, 2024, doi: 10.1080/01605682.2023.2210604.

[25] H. Rahmani and G. R. Weckman, "Working under the Shadow of Drones: Investigating Occupational Safety Hazards among Commercial Drone Pilots," *IISE Trans. Occup. Ergon. Hum. Factors*, vol. 12, no. 1–2, pp. 55–67, 2024, doi: 10.1080/24725838.2023.2251009.

[26] Y. A. Singgalen, "Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT)," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1551–1562, 2023, doi: 10.30865/mib.v7i3.6461.

[27] S. Leon, C. Chen, and A. Ratcliffe, "Consumers' perceptions of the environmental and public health benefits of last mile drone delivery," *Int. J. Logist. Res. Appl.*, vol. 0, no. 0, pp. 1–25, 2024, doi: 10.1080/13675567.2024.2341851.

[28] B. Alsadik, "Crowdsource Drone Imagery–A Powerful Source for the 3D Documentation of Cultural Heritage at Risk," *Int. J. Archit. Herit.*, vol. 16, no. 7, pp. 977–987, 2022, doi: 10.1080/15583058.2020.1853851.

[29] H. Han, "Consumer behavior and environmental sustainability in tourism and hospitality: a review of theories, concepts, and latest research," *J. Sustain. Tour.*, vol. 29, no. 7, pp. 1021–1042, 2021, doi: 10.1080/09669582.2021.1903019.

[30] G. Miller and A. Torres-Delgado, "Measuring sustainable tourism: a state of the art review of sustainable tourism indicators," *J. Sustain. Tour.*, vol. 31, no. 7, pp. 1483–1496, 2023, doi: 10.1080/09669582.2023.2213859.

[31] X. Zhan, F. bin Guo, and E. Roberts, "Image-based research methods for mapping tourist behaviour: smart photos," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–7, 2023, doi: 10.1080/02508281.2023.2196487.