Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

Sentiment and Toxicity Analysis of Sport Event MotoGP Mandalika Circuit Using Cross-Industry Standard Process for Data-Mining

Yerik Afrianto Singgalen

Faculty of Business Administration and Communication, Tourism Study Program, Atma Jaya Catholic University of Indonesia, Jakarta
Jl. Jend. Sudirman No.51 5, RT.004/RW.4, Karet Semanggi, Kecamatan Setiabudi, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta, Indonesia
Email: yerik.afrianto@atmajaya.ac.id
Correspondence Author Email: yerik.afrianto@atmajaya.ac.id
Submitted: **28/03/2024;** Accepted: **30/04/2024;** Published: **30/04/2024**

Abstract-This research identifies a research gap in understanding the impact of contextual factors on sentiment and toxicity within online discussions of sports events, focusing on the MotoGP event in Mandalika. By exploring how contextual nuances influence public sentiment and toxicity levels, this study aims to enhance the effectiveness of online discourse management and improve user experiences in digital platforms hosting event-related content. This research investigates the nuances of public sentiment and toxic language within textual data, specifically focusing on content videos of the MotoGP event held in Mandalika. Methodologically, the study embraces the CRISP-DM framework, facilitating structured data analysis, model development, and subsequent deployment. The findings reveal promising outcomes in terms of the performance of machine learning algorithms; notably, the k-NN algorithm attains an accuracy rate of 94.33%, precision of 96.48%, recall of 92.01%, f-measure of 94.19%, and an AUC score of 0.982. Similarly, the Support Vector Machine (SVM) demonstrates commendable accuracy, achieving 87.54%, precision of 99.53%, recall of 75.45%, f-measure of 85.82%, and an AUC score of 0.986. Furthermore, the toxicity analysis uncovers varying levels of harmful language, ranging from 0.01229 to 0.08933. These findings underscore the imperative nature of considering both sentiment dynamics and toxicity in managing online discourse effectively and enhancing user experiences across digital platforms. The sentiment analysis underscores the importance of understanding and effectively managing public emotions in the context of sports events like MotoGP. By acknowledging and addressing positive and negative sentiments, event organizers can better engage with their audience, mitigate potential issues, and ultimately enhance the overall experience for all involved.

Keywords: Sentiment; Sport Event; Tourism; Toxicity; SVM; DT; k-NN

1. INTRODUCTION

Sports events have emerged as integral components driving the growth of tourism economies worldwide. These events, characterized by competitive video gaming, draw massive crowds and generate substantial revenue streams. By attracting participants and spectators alike, sports tournaments bolster local economies through increased tourism spending on accommodations, dining, and entertainment [1]–[4]. Additionally, these events often showcase cutting-edge technological advancements, further enhancing the appeal of host destinations as innovative hubs [5]–[8]. Given their global popularity and capacity to attract diverse audiences, sports events undoubtedly play a pivotal role in stimulating economic growth within the tourism sector.

Integrating sport event MotoGP with tourism infrastructure presents a dynamic synergy that propels both industries forward. Sports event MotoGP tournaments mirror the exhilarating experience of real-life motorcycle racing, captivate audiences globally, drawing enthusiasts to host destinations [9]–[13]. As these events unfold virtually, they offer a unique opportunity for fans to engage with the sport more profoundly, transcending geographical barriers [14]–[16]. Moreover, hosting sports events and MotoGP competitions enhances the destination's profile and stimulates local economies through increased tourism expenditure [17], [18]. With its potential to attract diverse audiences and promote destination awareness, the collaboration between the sports event MotoGP and tourism underscores a mutually beneficial relationship driving economic growth and cultural exchange.

Incorporating the MotoGP sports event in Mandalika is a strategic initiative to revitalize Indonesia's economy through tourism. With Mandalika poised to become a hub for sports and MotoGP events, the region benefits from heightened international attention and increased visitor traffic [19]. As sports events and MotoGP competitions unfold in this picturesque locale, they not only showcase the natural beauty of Mandalika but also leverage the region's infrastructure developments, such as the Mandalika Circuit, to attract global audiences [20]. Furthermore, by positioning Mandalika as a premier Sports event MotoGP destination, Indonesia capitalizes on the event's popularity to drive economic recovery post-pandemic [21]–[24]. In conclusion, integrating the sports event MotoGP in Mandalika emerges as a strategic tool for leveraging Indonesia's tourism potential and catalyzing regional economic resurgence.

This research aims to analyze public sentiment regarding the organization of the MotoGP sports event in Mandalika through SPOTV Indonesia video content (ll-tMQvoaYo). By scrutinizing the discourse surrounding this event on online platforms, including social media and video-sharing platforms, this research gauges the attitudes and perceptions of the public towards the initiative. Through sentiment analysis techniques, the study seeks to identify prevalent themes, opinions, and sentiments expressed by the online community regarding the

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

MotoGP event in Mandalika. Such an investigation is crucial for understanding the public's reception of the event and informing future event organization and promotion strategies [25]–[28]. In conclusion, this research provides valuable insights into public sentiment surrounding the MotoGP event in Mandalika, contributing to a more comprehensive understanding of its impact and effectiveness in engaging audiences.

The method utilized in sentiment analysis is CRISP-DM (Cross-Industry Standard Process for Data Mining), a systematic and widely recognized framework for data analysis. This structured approach involves several vital stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. This research systematically navigates through these stages by employing CRISP-DM, ensuring a rigorous and comprehensive analysis of sentiment data related to the MotoGP event in Mandalika. This method offers a systematic framework for organizing and executing sentiment analysis tasks, allowing for efficient data processing, accurate modeling, and robust evaluation of results [29]–[31]. Thus, adopting CRISP-DM in sentiment analysis enhances the reliability and validity of findings, facilitating informed decision-making and actionable insights for stakeholders involved in event planning and management.

The urgency of this research lies in its potential to inform evidence-based decision-making and strategic planning in the context of organizing the MotoGP sports event in Mandalika. Understanding public sentiment is paramount, with the event poised to impact the local economy significantly, the tourism sector, and community engagement [32]–[34]. By promptly analyzing the sentiments expressed in online content, stakeholders adapt their approaches to address concerns, capitalize on positive perceptions, and enhance the event's overall success. Moreover, given the dynamic nature of online discourse, timely research is essential to capture evolving public opinions and sentiments, allowing for agile responses and effective stakeholder engagement [35]–[38]. In conclusion, the urgency of this research is underscored by its pivotal role in shaping the strategies and outcomes of the MotoGP event in Mandalika, ensuring its alignment with public expectations and maximizing its socioeconomic benefits.

This research's theoretical and practical implications are profound, offering valuable insights for academia and industry stakeholders. From a theoretical standpoint, this study contributes to the burgeoning field of sentiment analysis by applying advanced techniques to analyze public perceptions surrounding the MotoGP event in Mandalika. By exploring the intricacies of online discourse, this research enriches existing frameworks and methodologies for studying public sentiment in the context of large-scale sporting events. On a practical level, the findings of this research inform event organizers, policymakers, and marketing professionals about the effectiveness of their strategies in engaging and resonating with target audiences. By understanding the sentiments expressed by the public, stakeholders tailor their approaches to enhance event experiences, optimize promotional efforts, and foster positive community relations. Thus, this research's theoretical and practical implications underscore its significance in shaping scholarly discourse and real-world decision-making processes related to event organization and promotion.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis

Gap analysis is conducted to discern theoretical and practical opportunities for contributing to the nexus of sports events and tourism. This systematic process identifies disparities between current practices and desired outcomes, illuminating areas where improvements or innovations are made. By scrutinizing existing literature, industry practices, and stakeholder perspectives, gap analysis reveals the untapped potential and areas for growth within the realm of sports event tourism. Through this methodical examination, this research pinpoints theoretical gaps in scholarly understanding and practical gaps in industry practices, paving the way for targeted interventions and strategic initiatives. Ultimately, gap analysis is vital for harnessing the synergies between sports events and tourism, fostering sustainable development, and maximizing socioeconomic benefits for host destinations.

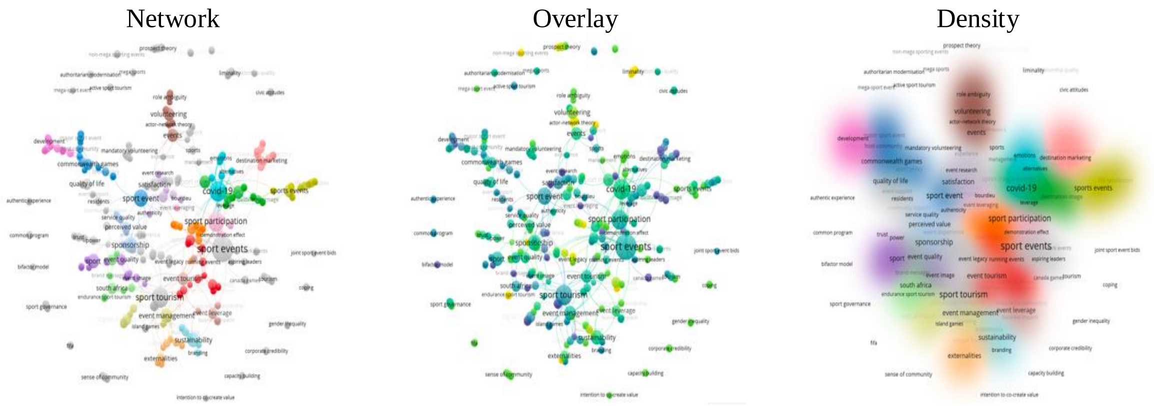


Figure 1. Network, Overlay, and Density Visualization using Vosviewer

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

Figure 1 shows the network, overlay, and density visualization using VosViewer. Based on the gap analysis of sports events and tourism results, it is evident that further examination of sentiment classification is warranted to analyze public sentiment regarding sports events in the tourism sector, with the MotoGP case serving as a pertinent example. By delving into sentiment classification, this research effectively gauges public perceptions and attitudes towards sports events, such as MotoGP, and their impact on tourism. This analytical approach enables a nuanced understanding of how public sentiment influences destination choices, visitor experiences, and marketing strategies [39], [40]. Thus, conducting sentiment classification studies within the context of sports events and tourism is crucial for informing decision-making processes and enhancing the overall management and promotion of such events to optimize their socioeconomic contributions to host destinations.

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

This research employs the CRISP-DM framework for sentiment analysis utilizing the k-NN, SVM, and DT algorithms. CRISP-DM provides a structured approach to data mining, encompassing various stages such as data understanding, data preparation, modeling, evaluation, and deployment. This research effectively manages the complexities of sentiment analysis tasks by adhering to this systematic framework, ensuring thoroughness and accuracy. Furthermore, the utilization of diverse algorithms, including k-NN (k-Nearest Neighbors), SVM (Support Vector Machine), and DT (Decision Trees), enhances the robustness and comprehensiveness of the analysis, enabling a comprehensive exploration of sentiment patterns and trends related to the MotoGP event and its impact on tourism. In conclusion, integrating the CRISP-DM framework and multiple algorithms in sentiment analysis underscores this research's methodological rigor and depth, contributing to a more nuanced understanding of public sentiment in the context of sports events and tourism.

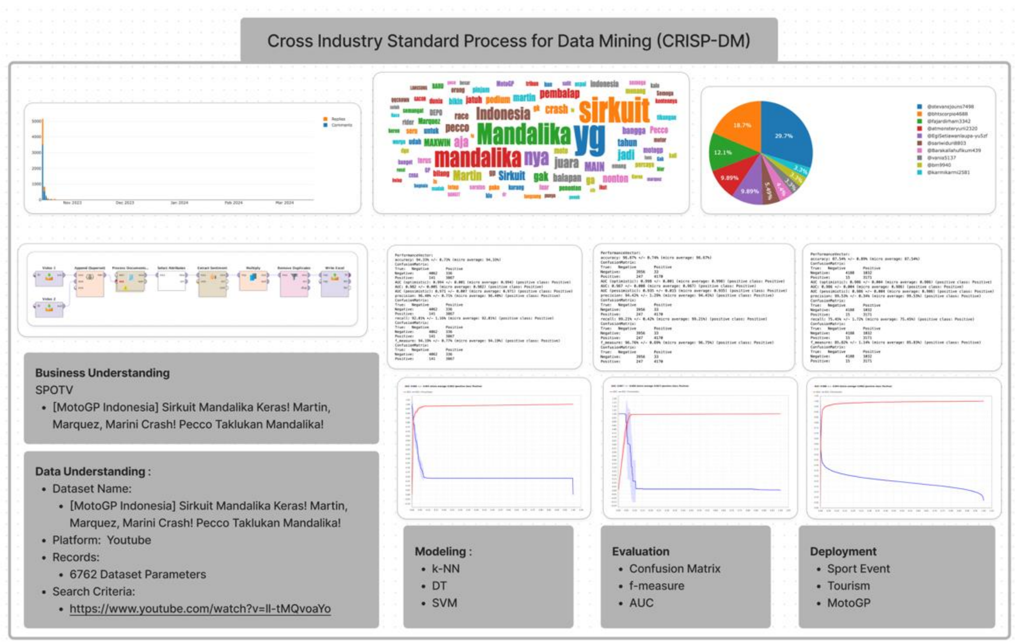


Figure 2. CRISP-DM Framework

Figure 2 shows the CRISP-DM framework. The CRISP-DM proves highly pertinent to the context of this research, particularly in the sentiment classification process employing k-NN, SVM, and DT models integrated with the SMOTE operator via the RapidMiner application. The CRISP-DM framework offers a structured methodology for navigating through the various stages of data mining, facilitating systematic analysis and interpretation of sentiment data related to the MotoGP event and its implications for tourism. By incorporating machine learning algorithms like k-NN, SVM, and DT, this research effectively classifies sentiments expressed in online content, providing valuable insights into public perceptions and attitudes toward sports events. Furthermore, integrating the SMOTE operator addresses class imbalance issues, enhancing the accuracy and reliability of sentiment analysis results. Thus, using CRISP-DM and advanced machine learning techniques underscores this research's methodological rigor and sophistication. This contributes to a comprehensive understanding of sentiment dynamics within the sports event and tourism domain.

2.2.1 Business Understanding

During the business understanding phase, the discussion context predominantly centers on the MotoGP sports event in Mandalika through SPOTV video content with the ID ll-tMQvoaYo. This initial phase of the CRISP-DM framework involves thoroughly comprehending the research endeavor's objectives, requirements, and constraints. By focusing on the specific context of the MotoGP event in Mandalika and the corresponding SPOTV video

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

content, this research clearly understands the key factors influencing public sentiment and perceptions surrounding this significant sporting event. Consequently, this phase lays a solid foundation for subsequent stages of the analysis, ensuring that the research remains aligned with its overarching goals and objectives.

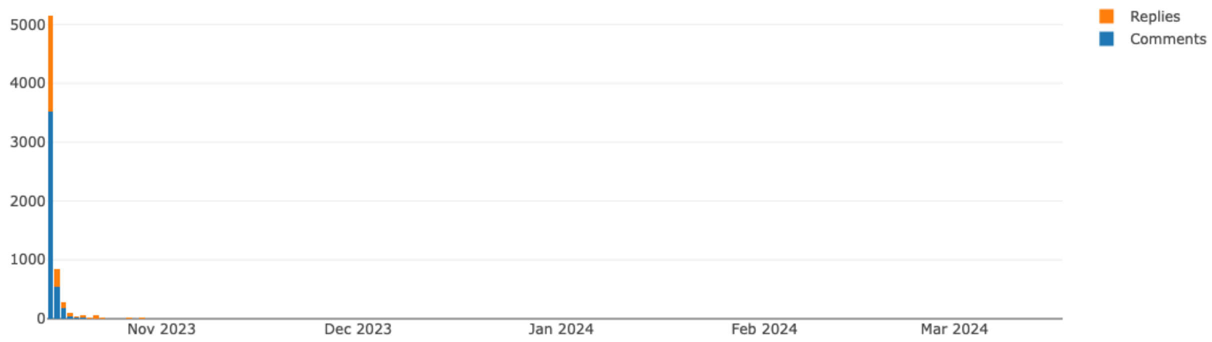


Figure 3. Post-per-day statistic of the Content (Communalytic)

Figure 3 shows the post-per-day statistic data of the content. Based on statistical data regarding post-per-day metrics, it is evident that on October 15, 2023, there were 3519 comments and 1640 replies, while on October 16, 2023, there were 553 comments and 300 replies. These figures underscore substantial engagement and interest in the content related to the MotoGP sports event in Mandalika. The surge in comments and replies across consecutive days signifies a heightened enthusiasm among online users for discussions of this significant sporting event, highlighting its relevance and appeal within the online community. Such robust engagement metrics reflect the considerable interest and anticipation surrounding the MotoGP event in Mandalika, suggesting its potential to captivate audiences and drive engagement on digital platforms.

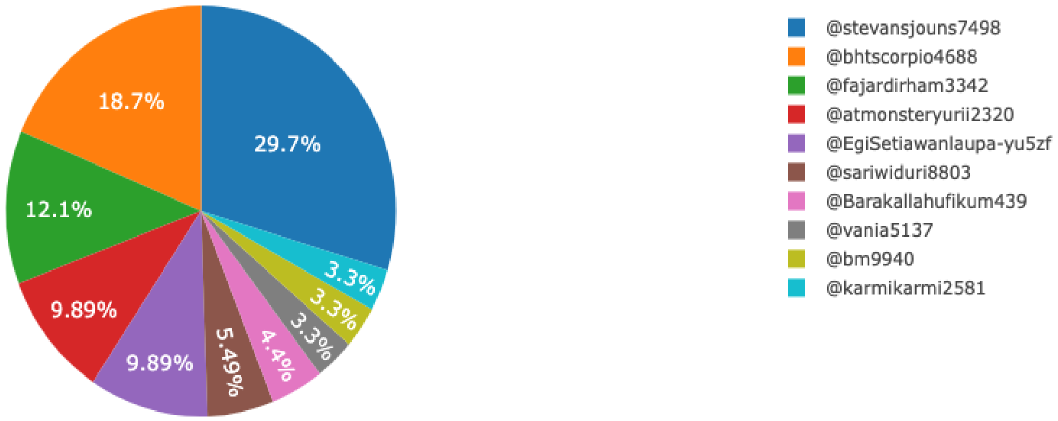


Figure 4. Top Ten Poster of the Content (Communalytic)

Figure 4 shows the top ten posters of the content. Based on statistical data of the top ten posters, it is observed that @stevansjouns7498 ranks highest with 27 posts, followed by @bhtscorpio4688 with 17 posts, and @fajardirham3342 with 11 posts, indicating varying levels of contribution and engagement within the online discourse surrounding the MotoGP event. The distribution of posts among these users suggests diverse perspectives and interests among active participants, enriching the dialogue and fostering a multifaceted discussion on this significant sporting event. This data underscores the dynamic nature of online interactions and the varied contributions made by different individuals to the discourse surrounding the MotoGP event, reflecting the complexity and depth of public engagement with the topic.

2.2.2 Data Understanding

During the data understanding phase, it becomes apparent that frequently used words play a crucial role in identifying significant topics that captivate content video reviewers. Analyzing the frequency and distribution of words within the dataset allows this research to discern recurring themes, key concepts, and prevalent sentiments viewers express. By identifying commonly used words and phrases, analysts gain insights into the overarching themes and subjects that resonate most strongly with the audience, guiding subsequent analysis and interpretation of the sentiment data. Therefore, leveraging frequently used words as a foundational element of data understanding facilitates more profound comprehension of the content video's appeal and relevance to viewers, enabling this research to extract meaningful insights and draw informed conclusions regarding public sentiment and engagement with the content.

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

Figure 5. Frequently Used Words (Communalytic)

Figure 5 shows the frequently used words. Specifically, the most frequently occurring words indicate vital themes and topics that resonate strongly within the discourse surrounding the MotoGP event in Mandalika. Words such as "sirkuit" (circuit) appeared 100 times, "Mandalika" 83 times, and "Indonesia" 47 times, reflecting the significance of the event's location and national context. Additionally, terms like "juara" (champion) were mentioned 39 times, "Martin" 37 times, and "crash" 33 times, highlighting the intensely competitive nature of MotoGP racing and the focus on race outcomes and performances. Furthermore, the prominence of names such as "pecco" with 31 mentions and "Marquez" with 18 mentions underscores the importance of specific individuals and their roles within the sport. These frequently used words serve as valuable indicators of the prevailing topics, sentiments, and interests driving discussions surrounding the MotoGP event, providing essential insights for further analysis and interpretation of public sentiment and engagement with the event.

Figure 6. Emoji Clouds (Communalytic)

Figure 6 shows the emoji clouds. Specifically, the most frequently occurring emojis shed light on the prevailing sentiments and reactions within discussions surrounding the MotoGP event in Mandalika. The 😂 emoji appeared 83 times, followed closely by the 👈 emoji with 77 occurrences and ❤ with 48 occurrences, indicating a mix of amusement, attention-grabbing, and affectionate sentiments expressed by participants. Additionally, emojis such as 👍 (33 occurrences), 😁 (31 occurrences), and 🎉 (23 occurrences) convey approval, lightheartedness, and celebratory responses to various aspects of the event. Furthermore, emojis like ID (23 occurrences) reflect national pride and patriotism among participants, particularly concerning Indonesia's hosting of the MotoGP event. These frequently used emojis serve as valuable indicators of the diverse emotions, reactions, and cultural nuances embedded within online discussions, enriching the understanding of public sentiment and engagement with the MotoGP event in Mandalika.

2.2.3 Modeling

During the modeling phase, the sentiment extraction process precedes evaluating the performance of k-NN, SVM, and DT algorithms in RapidMiner. This systematic approach ensures that sentiment analysis is conducted comprehensively and accurately before assessing the effectiveness of various machine-learning techniques in classifying sentiments. This research generates labeled datasets essential for training and testing the algorithms by extracting sentiment from the data. Subsequently, evaluating the performance of k-NN, SVM, and DT algorithms

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

allows for an informed comparison of their efficacy in sentiment classification, providing valuable insights into the most suitable approach for analyzing public sentiment related to the MotoGP event in Mandalika. Thus, conducting sentiment extraction before algorithm evaluation enhances the reliability and validity of the sentiment analysis process, facilitating more robust and insightful conclusions regarding public sentiment and engagement with the event.

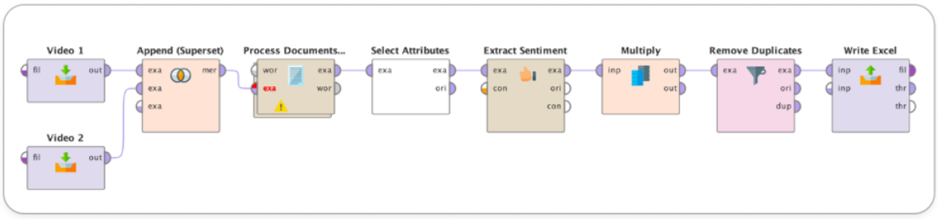


Figure 7. Extract Sentiment

Figure 7 shows the extract sentiment process in Rapidminer. Sentiment extraction aims to calculate negative and positive scores based on textual data. This essential step in sentiment analysis involves systematically analyzing the language used in texts to determine the prevailing sentiments expressed within them. By assigning numerical scores to words and phrases indicative of positive or negative sentiment, sentiment extraction enables quantitative assessment of the overall sentiment polarity of a given text. Through this process, derive insights into the emotional tone, attitudes, and opinions conveyed in textual content, facilitating a deeper understanding of public sentiment and engagement with the subject matter. Ultimately, sentiment extraction is crucial in enabling data-driven analysis and decision-making, contributing to more informed interpretations and strategies regarding public sentiment in contexts such as the MotoGP event in Mandalika.

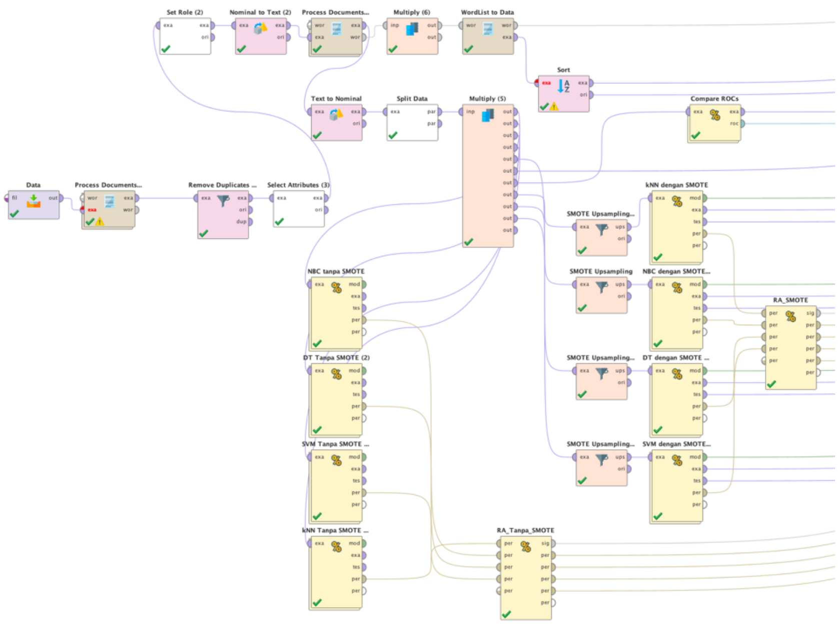


Figure 8. Modeling Process Using NBC, DT, k-NN, and SVM (Rapidminer)

Figure 8 shows the modeling process in Rapidminer. The extracted data is classified based on positive and negative sentiments before undergoing performance testing of NBC, k-NN, SVM, and DT algorithms. The findings of this research indicate that k-NN, SVM, and DT algorithms utilizing SMOTE demonstrate superior performance compared to others. This outcome underscores the effectiveness of advanced techniques such as SMOTE in mitigating class imbalance and enhancing sentiment classification accuracy. Leveraging k-NN, SVM, and DT algorithms in conjunction with SMOTE improves the reliability of sentiment analysis. It provides valuable insights for understanding public sentiment in contexts such as the MotoGP event in Mandalika.

2.2.4 Evaluation

Each algorithm or model is assessed based on confusion matrices, encompassing accuracy, precision, recall, f-measure, and AUC metrics during the evaluation stage. This comprehensive evaluation process enables this research to gauge the performance of the sentiment classification models across various dimensions, including their ability to correctly classify positive and negative sentiments, minimize misclassifications, and maintain overall predictive accuracy. By considering multiple evaluation metrics within the context of confusion matrices,

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

this research obtains a holistic understanding of the strengths and limitations of each algorithm or model, facilitating informed decision-making regarding their suitability for sentiment analysis tasks. Consequently, leveraging these evaluation techniques ensures robust and reliable assessments of sentiment classification performance, enhancing the validity and utility of research findings in analyzing public sentiment, particularly in domains such as the MotoGP event in Mandalika.

2.2.5 Deployment

The deployment stage of this research involves implementing and utilizing the developed sentiment analysis models and findings in real-world applications or decision-making processes. Through deployment, the insights garnered from the sentiment analysis of the MotoGP event in Mandalika inform various stakeholders, including event organizers, policymakers, and marketing professionals, in shaping strategies, interventions, and initiatives aimed at enhancing public engagement, optimizing event management practices, and maximizing the socioeconomic impact of the event. By operationalizing the findings of this research, stakeholders leverage actionable insights derived from sentiment analysis to guide their actions and decisions, ultimately contributing to the successful execution and outcomes of the MotoGP event in Mandalika. Therefore, the deployment stage is crucial in translating research findings into tangible outcomes and driving positive change in sports event management and tourism development.

3. RESULT AND DISCUSSION

Based on the outcomes of implementing CRISP-DM, each model exhibits distinct performance, namely k-NN, DT, and SVM. The CRISP-DM methodology enables a systematic approach to assessing the efficacy of various machine learning algorithms. Across the experimental iterations, it becomes evident that the k-NN, DT, and SVM models manifest divergent capabilities in handling the complexities of the dataset. Consequently, this underscores the significance of employing a comprehensive evaluation framework like CRISP-DM to discern the nuanced nuances of model performance, thereby facilitating informed decision-making in predictive modeling endeavors.

The discussion in this research centers on the performance of k-NN, DT, and SVM algorithms utilizing SMOTE. By focusing on these algorithms and incorporating SMOTE, this research addresses the challenge of class imbalance in sentiment analysis datasets, thereby enhancing the effectiveness and robustness of sentiment classification models. Through a systematic evaluation of the performance of these algorithms, this research gains insights into their efficacy in accurately classifying sentiments related to the MotoGP event in Mandalika. This focused discussion enables a thorough examination of the strengths and limitations of each algorithm, ultimately contributing to the advancement of sentiment analysis methodologies and the refinement of strategies for analyzing public sentiment in the context of sports events and tourism.

Performance Vector

```
PerformanceVector:
accuracy: 94.33% +/- 0.73% (micro average: 94.33%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       4062    336
Positive:       141     3867
AUC (optimistic): 0.994 +/- 0.001 (micro average: 0.994) (positive class: Positive)
AUC: 0.982 +/- 0.005 (micro average: 0.982) (positive class: Positive)
AUC (pessimistic): 0.971 +/- 0.007 (micro average: 0.971) (positive class: Positive
precision: 96.48% +/- 0.71% (micro average: 96.48%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       4062    336
Positive:       141     3867
recall: 92.01% +/- 1.16% (micro average: 92.01%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       4062    336
Positive:       141     3867
f_measure: 94.19% +/- 0.77% (micro average: 94.19%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       4062    336
Positive:       141     3867
```

Area Under Curve

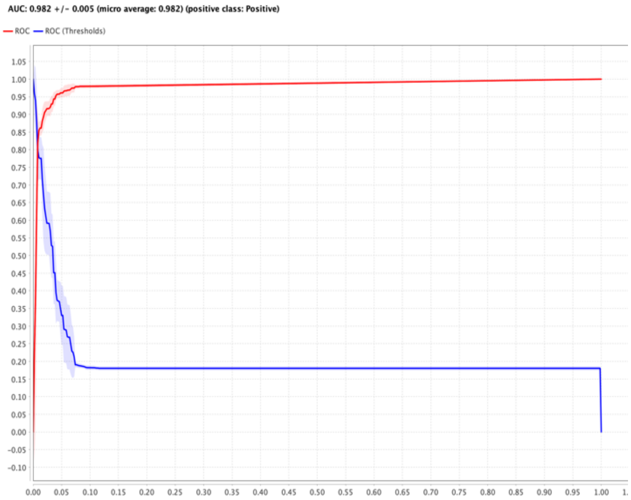
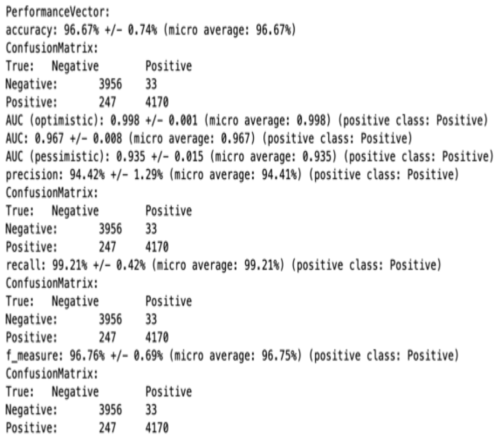


Figure 9. Performance of the k-NN Model with SMOTE in Sentiment Classification

Figure 9 shows the performance of the k-NN model with SMOTE in sentiment classification. Based on the evaluation of k-NN performance, it is evident that the algorithm exhibits exceptional accuracy, precision, recall, f-measure, and AUC values. With an accuracy rate of 94.33%, precision of 96.48%, recall of 92.01%, f-measure of 94.19%, and an AUC score of 0.982, the k-NN algorithm demonstrates its efficacy in accurately classifying sentiments related to the MotoGP event in Mandalika. These robust performance metrics underscore the reliability and effectiveness of the k-NN algorithm in sentiment analysis tasks, indicating its suitability for analyzing public sentiment and engagement with sports events.

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

Performance Vector



```
PerformanceVector:
accuracy: 96.67% +/- 0.74% (micro average: 96.67%)
ConfusionMatrix:
True:	Negative	Positive
Negative:	3956	33
Positive:	247	4170
AUC (optimistic): 0.998 +/- 0.001 (micro average: 0.998) (positive class: Positive)
AUC: 0.967 +/- 0.008 (micro average: 0.967) (positive class: Positive)
AUC (pessimistic): 0.935 +/- 0.015 (micro average: 0.935) (positive class: Positive)
precision: 94.42% +/- 1.29% (micro average: 94.41%) (positive class: Positive)
ConfusionMatrix:
True:	Negative	Positive
Negative:	3956	33
Positive:	247	4170
recall: 99.21% +/- 0.42% (micro average: 99.21%) (positive class: Positive)
ConfusionMatrix:
True:	Negative	Positive
Negative:	3956	33
Positive:	247	4170
f_measure: 96.76% +/- 0.69% (micro average: 96.75%) (positive class: Positive)
ConfusionMatrix:
True:	Negative	Positive
Negative:	3956	33
Positive:	247	4170
```

Area Under Curve

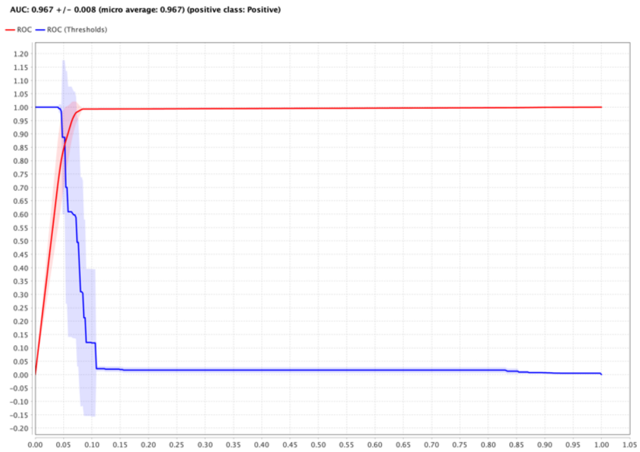
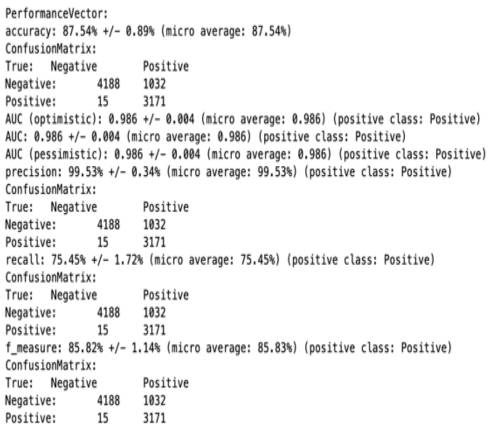


Figure 10. Performance of the DT Model with SMOTE in Sentiment Classification

Figure 10 shows the performance of the DT model with SMOTE in sentiment classification. Based on the evaluation of DT performance, it is evident that the algorithm demonstrates impressive accuracy, precision, recall, f-measure, and AUC values. With an accuracy rate of 96.67%, precision of 94.42%, recall of 99.21%, f-measure of 96.76%, and an AUC score of 0.967, the DT algorithm showcases its efficacy in accurately classifying sentiments related to the MotoGP event in Mandalika. These robust performance metrics underscore the reliability and effectiveness of the DT algorithm in sentiment analysis tasks, indicating its suitability for analyzing public sentiment and engagement with sports events.

Performance Vector



```
PerformanceVector:
accuracy: 87.54% +/- 0.89% (micro average: 87.54%)
ConfusionMatrix:
True:	Negative	Positive
Negative:	4188	1032
Positive:	15	3171
AUC (optimistic): 0.986 +/- 0.004 (micro average: 0.986) (positive class: Positive)
AUC: 0.986 +/- 0.004 (micro average: 0.986) (positive class: Positive)
AUC (pessimistic): 0.986 +/- 0.004 (micro average: 0.986) (positive class: Positive)
precision: 99.53% +/- 0.34% (micro average: 99.53%) (positive class: Positive)
ConfusionMatrix:
True:	Negative	Positive
Negative:	4188	1032
Positive:	15	3171
recall: 75.45% +/- 1.72% (micro average: 75.45%) (positive class: Positive)
ConfusionMatrix:
True:	Negative	Positive
Negative:	4188	1032
Positive:	15	3171
f_measure: 85.82% +/- 1.14% (micro average: 85.83%) (positive class: Positive)
ConfusionMatrix:
True:	Negative	Positive
Negative:	4188	1032
Positive:	15	3171
```

Area Under Curve

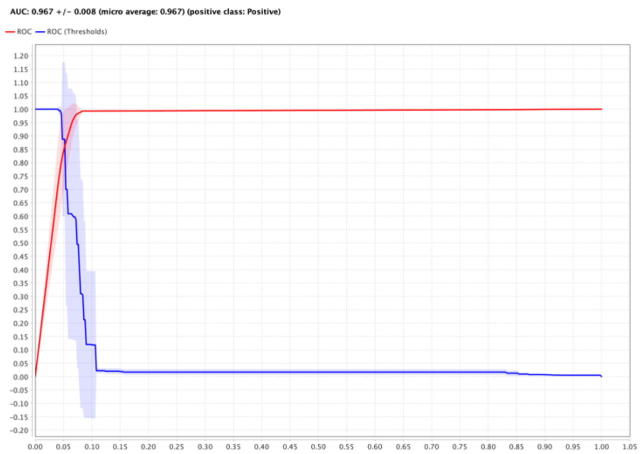


Figure 11. Performance of the SVM Model with SMOTE in Sentiment Classification

Figure 11 shows the performance of the SVM model with SMOTE in sentiment classification. Based on the evaluation of SVM performance, it is apparent that the algorithm demonstrates a commendable accuracy of 87.54%, precision of 99.53%, recall of 75.45%, f-measure of 85.82%, and AUC of 0.986. While the precision score indicates a high proportion of correctly classified positive sentiments among all predicted positive cases, the recall score signifies the algorithm's ability to identify a substantial portion of positive sentiments. However, the relatively lower recall score suggests the algorithm may miss some positive sentiments in the dataset. Nevertheless, the high accuracy, precision, and AUC values affirm the reliability and effectiveness of the SVM algorithm in accurately classifying sentiments related to the MotoGP event in Mandalika, indicating its potential as a valuable tool for sentiment analysis tasks.

The disparity in performance among the k-NN, DT, and SVM algorithms using SMOTE lies in their distinct approaches to handling class imbalance and their inherent characteristics in handling complex datasets. While k-NN relies on the proximity of data points for classification, DT utilizes a hierarchical decision tree structure to partition the feature space. On the other hand, SVM aims to find the optimal hyperplane that maximizes the margin between classes. Additionally, the effectiveness of SMOTE in addressing class imbalance may vary depending on the algorithm's sensitivity to the distribution of data points and the nature of the dataset. Therefore, the variation in performance highlights the importance of selecting the most suitable algorithm based on the specific characteristics and requirements of the sentiment analysis task, emphasizing the need for thorough experimentation and evaluation to determine the optimal approach.

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

In addition to sentiment classification, analyzing toxicity within content video review data is imperative. Toxicity analysis identifies and mitigates harmful or offensive language, attitudes, and behaviors in user-generated content. By scrutinizing the level of toxicity present in reviews, content creators, platform administrators, and policymakers take proactive measures to maintain a safe and respectful online environment. Moreover, toxicity analysis complements sentiment classification by providing a more nuanced understanding of the public's responses to content, facilitating more comprehensive insights into audience engagement and sentiment dynamics. Consequently, integrating toxicity analysis alongside sentiment classification enhances the effectiveness and relevance of data-driven approaches in understanding and managing online discourse surrounding content videos.

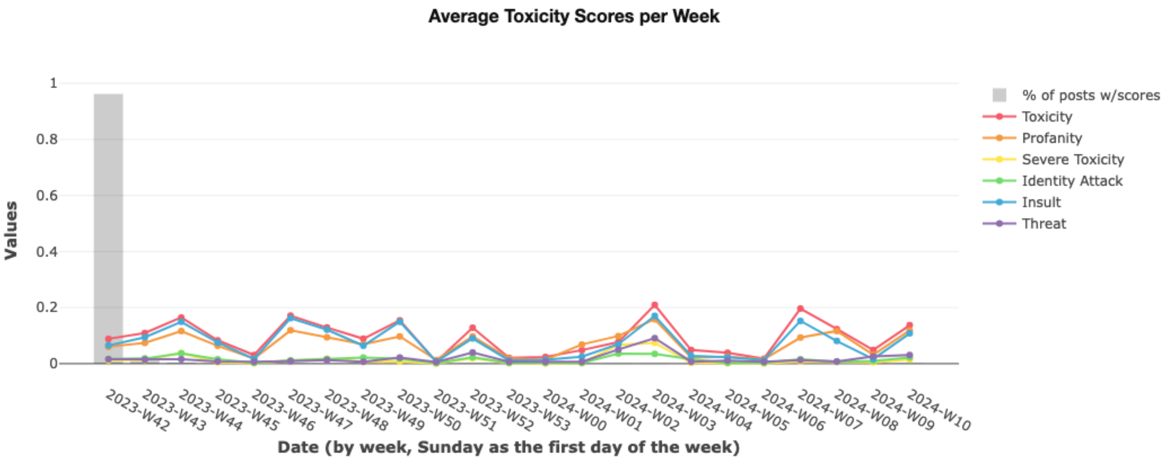


Figure 12. Average Toxicity Score per Week (Communalytic)

Figure 12 shows the average toxicity scores per week. Based on the results of the toxicity analysis, it is evident that the levels of toxicity across various categories vary significantly. With toxicity levels ranging from 0.01229 to 0.08933, the data reveals varying degrees of harmful or offensive language and behaviors within content video reviews. Additionally, the analysis indicates differences in the prevalence of severe toxicity, identity attacks, insults, profanity, and threats, with corresponding confidence scores ranging from 0.70217 to 0.99114. These findings underscore the importance of conducting comprehensive toxicity analysis to effectively identify and address problematic content, fostering a safer and more respectful online environment for users. Moreover, the nuanced insights provided by toxicity analysis complement other forms of content analysis, contributing to a more holistic understanding of user-generated content dynamics and facilitating informed decision-making by content creators, platform administrators, and policymakers.

Toxicity analysis provides insight into specific public sentiments regarding emotional relationships within the narrated text. By examining indicators such as severe toxicity, identity attacks, insults, profanity, and threats, this research discerns the emotional tone and interpersonal dynamics conveyed within textual content. This comprehensive understanding of public sentiment enables a nuanced exploration of emotional responses and attitudes expressed in user-generated narratives, enriching the interpretation of audience engagement and sentiment dynamics. Consequently, integrating toxicity analysis alongside other sentiment analysis methodologies enhances the depth and accuracy of textual data analysis, facilitating a more nuanced comprehension of public sentiment and emotional nuances within narrative texts.

4. CONCLUSION

In conclusion, the research findings underscore the significance of employing advanced analytical techniques, such as sentiment and toxicity analysis, in understanding public sentiment and emotional dynamics within textual data. The research adopts the CRISP-DM framework, emphasizing the structured approach to data mining tasks, from business understanding to deployment. Incorporating CRISP-DM facilitates systematic data analysis and model development, enhancing the research's rigor and efficiency. The analysis highlights the importance of considering both sentiment and toxicity aspects to comprehensively understand public sentiment and emotional nuances. By integrating these analytical approaches, researchers, content creators, and platform administrators make informed decisions to foster a safer, more respectful online environment and enhance user engagement with content. Through evaluating sentiment classification models and toxicity indicators, the study has provided valuable insights into the emotional responses, attitudes, and behaviors expressed in user-generated content, particularly in the context of content videos. With an accuracy rate of 94.33%, precision of 96.48%, recall of 92.01%, f-measure of 94.19%, and an AUC score of 0.982, the k-NN algorithm demonstrates its efficacy in accurately classifying sentiments related to the MotoGP event in Mandalika. Similarly, based on the evaluation of SVM performance, it is apparent that the algorithm demonstrates a commendable accuracy of 87.54%, precision of 99.53%, recall of 75.45%, f-measure of 85.82%, and AUC of 0.986. With toxicity levels ranging from 0.01229 to 0.08933, the data

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

reveals varying degrees of harmful or offensive language and behaviors within content video reviews. The analysis indicates differences in the prevalence of severe toxicity, identity attacks, insults, profanity, and threats, with corresponding confidence scores ranging from 0.70217 to 0.99114. Thus, the research contributes to advancing sentiment analysis and toxicity detection methodologies, ultimately facilitating more effective strategies for managing online discourse and enhancing user experiences in digital platforms. The findings significantly contribute to the field, guiding future research endeavors in analyzing public sentiment with a quantitative approach.

ACKNOWLEDGMENT

I sincerely thank the Tourism Department, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, and the Research and Community Service Institute (LPPM) for their invaluable support in facilitating the publication of this research.

REFERENCES

[1] P. N. Acha-Anyi, “Sports events and community development: Analysis of the AFCON 2022 host cities in Cameroon,” Cogent Soc. Sci., vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2195079.
[2] R. Ladhari and N. Souiden, “The role of mega-sports event experience and host city experience in explaining enjoyment, city image, and behavioral intentions,” J. Travel Tour. Mark., vol. 37, no. 4, pp. 460–478, 2020, doi: 10.1080/10548408.2020.1783427.
[3] Z. Maditinos, C. Vassiliadis, Y. Tzavlopoulos, and S. A. Vassiliadis, “Sports events and the COVID-19 pandemic: assessing runners’ intentions for future participation in running events–evidence from Greece,” Tour. Recreat. Res., vol. 46, no. 2, pp. 276–287, 2021, doi: 10.1080/02508281.2020.1847422.
[4] E. Pereira, M. Mascarenhas, A. Flores, L. Chalip, and G. Pires, “Strategic leveraging: evidences of small-scale sport events,” Int. J. Event Festiv. Manag., vol. 11, no. 1, pp. 69–88, Jan. 2020, doi: 10.1108/IJEFM-07-2018-0046.
[5] M. Johnston, M. Naylor, G. Dickson, and T. Kellison, “Insider Perspectives of a Major Sport Event Referendum,” Int. J. Sport Policy Polit., vol. 13, no. 4, pp. 605–622, 2021, doi: 10.1080/19406940.2021.1929405.
[6] E. Daigo and K. Filo, “Exploring the value sponsors co-create at a charity sport event: a multiple stakeholder perspective of sport value,” Sport Manag. Rev., vol. 25, no. 4, pp. 656–678, 2022, doi: 10.1080/14413523.2021.1975401.
[7] K. Filo, N. Hookway, M. Wade, and C. Palmer, “An exploration of charity sport event donor perceptions of online peer-to-peer fundraising mechanisms,” Sport Manag. Rev., vol. 25, no. 5, pp. 847–870, 2022, doi: 10.1080/14413523.2021.1993645.
[8] R. Zhou, K. Kaplanidou, and C. Wegner, “Social capital from sport event participation: scale development and validation,” Leis. Stud., vol. 40, no. 5, pp. 612–627, 2021, doi: 10.1080/02614367.2021.1916832.
[9] S. Owen and D. Chambers, “Volunteers’ Sense of (Dis)Connection at a Sport Event,” Leis. Sci., vol. 46, no. 2, pp. 105–122, 2024, doi: 10.1080/01490400.2021.1916660.
[10] K. Kaplanidou, S. F. Fleshman, and I. Cho, “Sport event travel intentions during times of crisis: the role of life goals, risk and emotions,” J. Sport Tour., vol. 27, no. 2, pp. 123–137, 2023, doi: 10.1080/14775085.2023.2186927.
[11] E. L. Lachance, A. Thompson, J. T. Bakhsh, and M. M. Parent, “Volunteer retention: Examining intentions and behaviours in the wrap-Up mode of a professional recurring small-scale sport event,” Manag. Sport Leis., pp. 1–17, 2022, doi: 10.1080/23750472.2022.2147859.
[12] J. Byun, D. Ellis, and B. Leopkey, “The pursuit of legitimacy through strategic alliances: the examination of international joint sport event bidding,” Eur. Sport Manag. Q., vol. 21, no. 4, pp. 544–563, 2021, doi: 10.1080/16184742.2020.1759668.
[13] B. L. Newland, T. J. Aicher, M. Davies, and E. Hungenberg, “Sport event ecotourism: sustainability of trail racing events in US National Parks,” J. Sport Tour., vol. 25, no. 2, pp. 155–181, 2021, doi: 10.1080/14775085.2021.1902374.
[14] S. Kim and A. E. Manoli, “Does relationship quality matter in policy-making? The impact of government-public relationships and residents’ perceptions on their support towards a mega-sport event,” Int. J. Sport Policy Polit., vol. 14, no. 2, pp. 207–224, 2022, doi: 10.1080/19406940.2021.2013925.
[15] K. Bodin and M. Taks, “Unpacking the public/government relationship in the context of sport events: an agency theory approach,” Int. J. Sport Policy Polit., vol. 14, no. 4, pp. 657–671, 2022, doi: 10.1080/19406940.2022.2102669.
[16] B. An, M. Harada, and S. Sato, “Service quality, satisfaction, and behavioral intention in a triathlon event: the different experiences between local and non-local participants,” J. Sport Tour., vol. 24, no. 2, pp. 127–142, 2020, doi: 10.1080/14775085.2020.1773296.
[17] J. Đurkin Badurina, M. Perić, and V. Vitezić, “Potential for the regeneration of rural areas through local involvement in the organisation of sport events,” Manag. Sport Leis., vol. 26, no. 5, pp. 377–394, 2021, doi: 10.1080/23750472.2020.1829990.
[18] T. M. Hayduk, “Do the rich get richer? Exploring disparate effects of hosting sport mega events on high technology exports for developed and developing nations,” J. Int. Trade Econ. Dev., vol. 29, no. 8, pp. 973–994, 2020, doi: 10.1080/09638199.2020.1782973.
[19] R. E. Caraka et al., “Connectivity, sport events, and tourism development of Mandalika’s special economic zone: A perspective from big data cognitive analytics,” Cogent Bus. Manag., vol. 10, no. 1, 2023, doi: 10.1080/23311975.2023.2183565.
[20] S. Y. C. Ng, A. Bloom, S. L. Corcoran, T. Fletcher, and J. Sibley, “‘If you respect us..listen to us’: how sporting event media reframes or reinforces representations of street-connected children,” Leis. Stud., vol. 41, no. 6, pp. 757–774, 2022, doi: 10.1080/02614367.2022.2088830.

Journal of Information System Research (JOSH)
Volume 5, No. 3, April 2024, pp 731−741
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v5i3.5056

[21] D. Svensson and A. Radmann, "Keeping Distance? Adaptation Strategies to the Covid-19 Pandemic Among Sport Event Organizers in Sweden," J. Glob. Sport Manag., vol. 8, no. 3, pp. 594–611, 2023, doi: 10.1080/24704067.2021.1936592.
[22] J. Rookwood, "From sport-for-development to sports mega-events: conflict, authoritarian modernisation and statecraft in Azerbaijan," Sport Soc., vol. 25, no. 4, pp. 847–866, 2022, doi: 10.1080/17430437.2021.2019710.
[23] K. Kaplanidou, A. Apostolopoulou, and I. Cho, "Sport Consumption Intentions during a Crisis: The COVID-19 Pandemic," J. Glob. Sport Manag., vol. 8, no. 4, pp. 985–1007, 2023, doi: 10.1080/24704067.2021.1991831.
[24] K. H. Choi, S. I. Choi, and J. Kim, "The influence of prospective event spectators' risk-taking tendency on COVID-19 risk perception and information-seeking: the case of the Tokyo 2020 Olympic Games," J. Sport Tour., vol. 27, no. 3, pp. 221–238, 2023, doi: 10.1080/14775085.2023.2201259.
[25] R. Bjerke and H. E. Naess, "Toward a co-Creation framework for developing a green sports event brand: the case of the 2018 Zürich E Prix," J. Sport Tour., vol. 25, no. 2, pp. 129–154, 2021, doi: 10.1080/14775085.2021.1895872.
[26] G. Cuskelly, L. Fredline, E. Kim, S. Barry, and P. Kappelides, "Volunteer Selection at a Major Sport Event: A Strategic Human Resource Management Approach," Sport Manag. Rev., vol. 24, no. 1, pp. 116–133, 2021, doi: 10.1016/j.smr.2020.02.002.
[27] A. Morgan, T. Taylor, and D. Adair, "Sport event sponsorship management from the sponsee's perspective," Sport Manag. Rev., vol. 23, no. 5, pp. 838–851, 2020, doi: 10.1016/j.smr.2020.04.006.
[28] G. B. Williams, Y. H. Kim, and J. Nauright, "Destination development by sport event tourism (SET): a case study of Thailand," Sport Soc., vol. 24, no. 10, pp. 1827–1837, 2021, doi: 10.1080/17430437.2021.1916234.
[29] Y. A. Singgalen, "Selling vegetables through live streaming : sentiment and network analysis," Int. J. Soc. Sci. Econ. Art, vol. 13, no. 4, pp. 240–254, 2024.
[30] Y. A. Singgalen, "Social network and sentiment analysis of product reviews (case of smartwatch product content)," Int. J. Soc. Sci. Econ. Art, vol. 13, no. 4, pp. 255–267, 2024.
[31] Y. A. Singgalen, "Toxicity and Social Network Analysis of Green Marketing Content for Electric Cars through Digital Media," Int. J. Soc. Sci. Econ. Art, vol. 13, no. 4, pp. 268–280, 2024.
[32] M. Johnston, M. Naylor, G. Dickson, D. Hedlund, and T. Kellison, "Determinants of support and participation in a major sport event referendum," Sport Manag. Rev., vol. 24, no. 1, pp. 134–155, 2021, doi: 10.1016/j.smr.2020.08.001.
[33] M. Duignan, J. Carlini, and M. Parent, "Host community salience loss across major sport event planning," Eur. Sport Manag. Q., vol. 0, no. 0, pp. 1–23, 2023, doi: 10.1080/16184742.2023.2237063.
[34] B. E. Menaker, D. Sheptak, J. Kurland, and D. Tekin, "Rethinking Sport Event Security: From Risk Management to a Community Driven Approach," J. Glob. Sport Manag., vol. 0, no. 0, pp. 1–23, 2021, doi: 10.1080/24704067.2021.1929388.
[35] K. Aizawa, M. Orr, Y. Inoue, J. Nagazumi, and M. Yoshida, "Leveraging sport events for sustainable sport participation: how schools contribute to sport development through events," Eur. Sport Manag. Q., vol. 23, no. 3, pp. 662–682, 2023, doi: 10.1080/16184742.2021.1910326.
[36] J. Hemmonsbey and T. M. Tichaawa, "Brand messages that influence the sport tourism experience: the case of South Africa," J. Sport Tour., vol. 24, no. 3, pp. 177–194, 2020, doi: 10.1080/14775085.2020.1822200.
[37] D. Cook, R. Biscaia, K. Papadas, L. Simkin, and L. Carter, "The creation of shared value in the major sport event ecosystem: understanding the role of sponsors and hosts," Eur. Sport Manag. Q., vol. 23, no. 3, pp. 811–832, 2023, doi: 10.1080/16184742.2021.1931394.
[38] D. Won, W. Chiu, C. Lee, and H. Bang, "The Impact of Internal Marketing Activities on Mandatory Volunteers in Sport: A Case from the 2019 Military World Games in China," J. Glob. Sport Manag., vol. 0, no. 0, pp. 1–21, 2023, doi: 10.1080/24704067.2023.2197460.
[39] V. Ziakas, "Leveraging Sport Events for Tourism Development: The Event Portfolio Perspective," J. Glob. Sport Manag., vol. 8, no. 1, pp. 43–72, 2023, doi: 10.1080/24704067.2020.1731700.
[40] T. Ströbel and C. C. Germelmann, "Exploring new routes within brand research in sport management: directions and methodological approaches," Eur. Sport Manag. Q., vol. 20, no. 1, pp. 1–9, 2020, doi: 10.1080/16184742.2019.1706603.