Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

Sentiment Classification of Climate Change and Tourism Content Using Support Vector Machine

Yerik Afrianto Singgalen

Faculty of Business Administration and Communication, Tourism Study Program, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: yerik.afrianto@atmajaya.ac.id
Submitted: 04/02/2024; Accepted: 20/02/2024; Published: 21/02/2024

Abstract−This research aims to classify public sentiment regarding the issue of climate change and tourism. The research problem addressed in this study pertains to the classification of public sentiment concerning climate change within the tourism sector. Specifically, the study aims to explore and classify the public's sentiments regarding the impact of climate change on tourism activities.The methodology employed is CRISP-DM, which encompasses stages of business understanding, data understanding, modeling, evaluation, and deployment. Specifically, the SVM and SMOTE algorithms are utilized in the modeling stage to achieve optimal results. By leveraging this systematic approach and advanced algorithms, the study seeks to comprehensively analyze public sentiment towards climate change within the context of tourism, thus contributing valuable insights to academia and industry practitioners. Applying CRISP-DM methodology coupled with SVM and SMOTE algorithms enhances the rigor and effectiveness of sentiment analysis in addressing the complexities of climate change discourse in the tourism sector. The findings of this research demonstrate that the SVM and SMOTE algorithms yield promising results in sentiment classification, with an accuracy of 86.15% +/- 1.68% (micro average: 86.15%), precision of 85.17% +/- 2.16% (micro average: 85.11%) (positive class: Positive), recall of 87.64% +/- 3.39% (micro average: 87.64%) (positive class: Positive), f_measure of 86.34% +/- 1.79% (micro average: 86.35%) (positive class: Positive), and AUC of 0.923 +/- 0.012 (micro average: 0.923) (positive class: Positive). These metrics indicate the effectiveness and reliability of the SVM and SMOTE algorithms in accurately classifying sentiment toward climate change in the context of tourism. The high accuracy, precision, recall, f_measure, and AUC scores suggest that the models produced by these algorithms are robust and capable of capturing nuanced sentiment patterns, thereby contributing to the advancement of sentiment analysis techniques in climate change research within the tourism domain.

Keywords: Sentiment; Classification; Climate Change; Tourism; Support Vector Machine

1. INTRODUCTION

The impact of climate change on tourism has emerged as a pivotal subject warranting comprehensive scrutiny. The overarching issue lies in its profound alterations on natural landscapes and ecosystems, directly affecting tourist destinations worldwide [1]. Furthermore, the escalating frequency and intensity of extreme weather events exacerbate these changes, posing significant challenges to the sustainability and resilience of tourism activities [2]. Consequently, there is an urgent need for thorough investigation and analysis to understand the multifaceted implications of climate change on tourism [3]. Such an endeavor is imperative for devising effective mitigation and adaptation strategies to safeguard the future viability of the tourism industry and the socio-economic well-being of communities reliant on it [4].

The issue of climate change has pervaded various digital media platforms, offering widespread accessibility to users of digital applications. This proliferation serves as a digital campaign to heighten awareness regarding climate change and its repercussions on human life [5]. As information dissemination becomes increasingly digitized, it facilitates the dissemination of knowledge and empowers individuals to engage with climate-related issues more actively [6]. Consequently, this digital campaign is crucial in fostering a collective understanding of the urgency to address climate change [7]. It underscores the importance of concerted efforts toward mitigating its adverse effects.

The issue of climate change exerts a profound influence on coastal and maritime tourist destinations, significantly impacting the sustainability of local tourism. The primary concern lies in the escalating sea levels, changing weather patterns, and the increasing frequency of extreme events, collectively threatening coastal areas' ecological integrity and aesthetic appeal [8]. These impacts, in turn, jeopardize maritime tourism's overall allure and economic viability [9]. Consequently, it is imperative to recognize the intricate interplay between climate change and coastal tourism, emphasizing the urgency of implementing sustainable practices to mitigate environmental degradation and ensure the long-term resilience of local tourism economies [10]. Given these challenges, a comprehensive and strategic approach is required to safeguard these valuable destinations, promoting a harmonious coexistence between tourism development and environmental conservation.

The inception of this research emanates from the imperative to address the multifaceted challenges associated with sentiment classification in the context of climate change and its implications for the tourism industry. The growing significance of sustainable tourism and the need to comprehend public perceptions on climate change impact have underscored the necessity for a nuanced sentiment analysis approach. The limited extant research on sentiment classification within the tourism domain prompted this investigation to fill the existing gap in scholarly discourse. The complexities of sentiment patterns related to climate change necessitated a comprehensive examination, prompting the adoption of advanced methodologies such as CRISP-DM, along with the utilization of SVM and SMOTE algorithms for optimal classification outcomes. In conclusion, the research is propelled by the exigency to contribute substantively to the understanding of public sentiment regarding climate change within the pivotal domain of tourism.

The urgency of this research underscores the significance of digital climate change campaigns and their impact on tourism, particularly in shaping public responses to the dangers of climate change and the comfort of tourist experiences. Digital campaigns are vital tools for disseminating information, raising awareness, and mobilizing public opinion on pressing environmental issues such as climate change [11]. By leveraging digital platforms, these campaigns can effectively reach diverse audiences, fostering a widespread understanding of the threats posed by climate change to tourist destinations [12]. Moreover, they influence tourists' perceptions and behaviors, encourage environmentally responsible practices, and promote sustainable tourism [13]. Thus, conducting research in this area is essential for elucidating the efficacy of digital campaigns in shaping public attitudes towards climate change and fostering sustainable tourism practices, ultimately contributing to the preservation of natural resources and the long-term viability of tourism industries worldwide.

The practical implications of this research emphasize the importance of collective public action in response to the issue of climate change through digital campaigns aimed at enhancing tourists' and local communities' understanding of anticipatory measures against climate change-induced disasters in tourism. By harnessing digital platforms for advocacy and information dissemination, these campaigns have the potential to mobilize widespread support and engagement, fostering a shared sense of responsibility toward mitigating the adverse impacts of climate change on tourist destinations [14]. Moreover, they facilitate the communication of proactive measures and adaptation strategies, empowering tourists and residents to make informed decisions and take necessary actions to safeguard the resilience and sustainability of tourism in the face of climate-related challenges [15]. In essence, this research underscores the pivotal role of digital campaigns in promoting collective action and building resilience within the tourism sector, thereby contributing to the long-term preservation of natural and cultural heritage for future generations.

The theoretical implications of this study underscore the significance of collective awareness regarding the impacts of climate change on livelihoods and ecologies in tourist destinations. The research contributes to the broader understanding of the interconnectedness between environmental sustainability, socio-economic well-being, and tourism development by delving into this aspect [16]. The study illuminates the intricate dynamics at play through theoretical exploration and empirical analysis, shedding light on how climate change-induced disruptions can reverberate throughout local communities and ecosystems reliant on tourism activities [17]. This deeper understanding enriches theoretical frameworks in tourism studies and informs policy-making and practical interventions to foster resilience and adaptive capacity within destination communities [18]. In conclusion, the theoretical implications of this research underscore the imperative of addressing climate change as an integral component of sustainable tourism development, thereby safeguarding the livelihoods and ecologies that underpin the viability of tourist destinations.

The limitation of this research lies in the methodology employed, precisely the employment of sentiment classification utilizing the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach and the Support Vector Machine (SVM) algorithm with the Synthetic Minority Over-Sampling Technique (SMOTE) operator [19]–[26]. While these methods offer valuable insights into understanding public sentiment towards climate change campaigns in tourism, they also pose certain constraints. Although widely used in data mining, the CRISP-DM approach may overlook nuances and contextual complexities inherent in sentiment analysis. Additionally, while SVM with SMOTE effectively handles imbalanced datasets, its reliance on predetermined parameters and assumptions may limit the generalizability of findings [27]. Therefore, future research endeavors should explore alternative methodologies or integrate complementary approaches to address these limitations comprehensively.

Further research recommendations should focus on comparing algorithms for sentiment classification of climate change content and its impacts on the tourism industry. This comparison would involve assessing the efficacy and accuracy of different algorithms, such as Support Vector Machine (SVM), Naive Bayes, Decision Trees, and Neural Networks, in analyzing sentiments expressed in digital content related to climate change and tourism. By conducting comparative analyses, researchers can identify the most suitable algorithm or combination of algorithms for accurately classifying sentiment about climate change and tourism in diverse datasets. Such investigations would enhance methodological rigor and contribute to developing robust analytical frameworks for understanding public perceptions and responses to climate change within the tourism context, thereby informing more effective strategies for sustainable tourism development.

2. RESEARCH METHODOLOGY

2.1 Research Gap and Trends Mapping: Climate Change and Tourism

In the process of identifying research gaps, this study utilized VOSviewer as a tool to map journals focusing on the topics of climate change and tourism. VOSviewer, a powerful bibliometric software, enables researchers to visually represent and analyze the relationships between scholarly publications based on citation patterns,

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

keywords, and co-authorship networks. By employing VOSviewer, the study effectively identified vital thematic areas, emerging trends, and potential gaps within the existing literature on climate change and tourism. This methodological approach provides valuable insights into the current state of research and facilitates identifying areas that warrant further investigation and exploration. Thus, leveraging VOSviewer in research endeavors contributes to a more systematic and comprehensive understanding of the research landscape, ultimately enhancing the quality and relevance of scholarly contributions in the field.

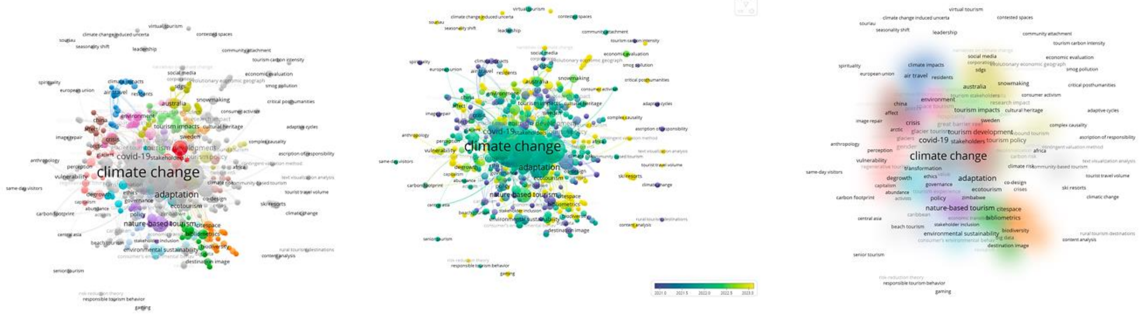

Figure 1. Network, Overlay, and Density Visualization

Based on the results of the research gap identification, it is evident that studies focusing on sentiment analysis of climate change and tourism content are relatively scarce, underscoring the importance of conducting comprehensive investigations utilizing the CRISP-DM methodology alongside the SVM and SMOTE algorithms. The limited attention to this specific intersection between climate change and tourism highlights a significant gap in the existing literature, necessitating a more thorough examination to elucidate public perceptions and responses in this domain. Researchers can gain deeper insights into sentiment patterns by employing robust methodologies such as CRISP-DM and leveraging advanced algorithms like SVM and SMOTE, thus enriching our understanding of the complex interactions between climate change discourse and tourism practices. In conclusion, addressing this research gap holds considerable potential for advancing knowledge and informing evidence-based strategies to foster sustainable tourism development in the face of climate change challenges.

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

The method employed in this research is the CRISP-DM, which comprises business understanding, data understanding, modeling, evaluation, and deployment stages. CRISP-DM offers a systematic framework for conducting data mining projects, guiding researchers through essential steps from understanding the business objectives to deploying the final model. By adhering to the CRISP-DM methodology, the research ensures a structured and comprehensive approach to analyzing sentiment toward climate change and tourism. This methodological rigor enhances the reliability and validity of the study's findings, contributing to a robust understanding of public sentiment in this critical area of inquiry.

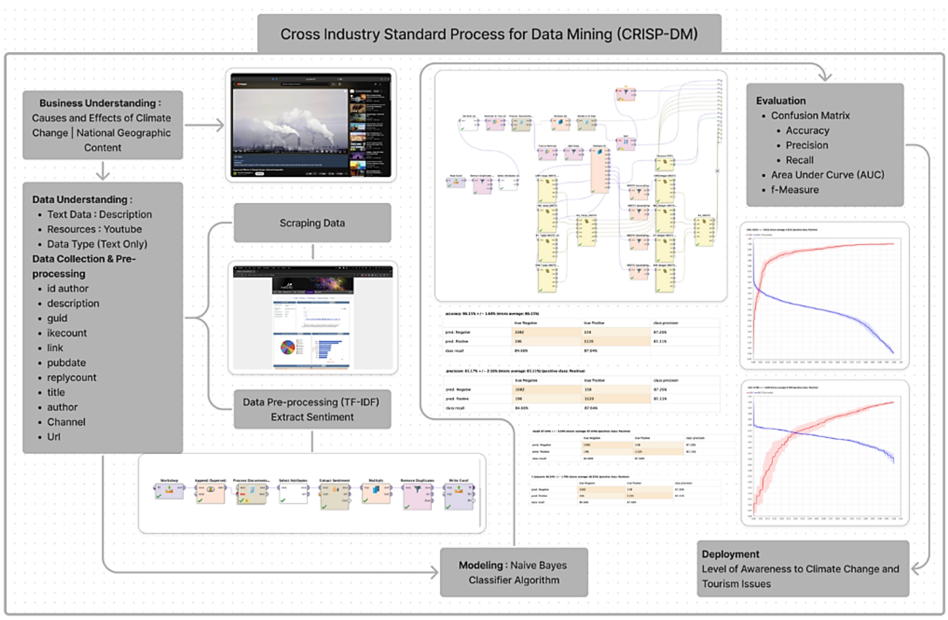

Figure 1. Implementaiton of CRISP-DM and SVM Algorithm

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

In the business understanding stage, the research utilizes videos from YouTube, focusing on climate change, as the primary data. In the data understanding stage, user reviews from the YouTube platform are employed as textual data to be classified and analyzed based on positive and negative sentiments. Subsequently, the SVM algorithm and SMOTE operator are employed in the modeling stage to generate a confusion matrix. This structured approach ensures a comprehensive understanding of public sentiment towards climate change conveyed through digital media, facilitating rigorous analysis and interpretation of the research findings. In the evaluation stage, the confusion matrix values concerning accuracy, precision, recall, Area Under Curve (AUC), and f-measure are assessed to measure the best performance of the SVM algorithm. Meanwhile, SMOTE addresses the issue of data imbalance that impacts the algorithm's performance. This comprehensive evaluation ensures the robustness and reliability of the sentiment analysis outcomes, providing valuable insights into public perceptions of climate change within the tourism context.

In the data understanding phase, textual data was sourced from public comments on video content published by National Geographic, identified by the video ID G4H1N_yXBiA, comprising a total of 3495 comments. The selection of this specific dataset is grounded in the richness and diversity of insights that public comments on National Geographic videos offer, reflecting a broad spectrum of perspectives on climate change within the context of the content disseminated by a renowned scientific and educational platform. Analyzing this dataset provides a unique opportunity to glean valuable insights into public sentiment, contributing to a more comprehensive understanding of the nuanced discourse surrounding climate change as presented by National Geographic. In conclusion, the chosen dataset aligns with the research objective, allowing for an in-depth exploration of sentiment patterns within a contextually relevant and substantial pool of public commentary on climate change.

During the modeling phase, a dataset consisting of 3495 textual comments was extracted using the Extract Sentiment function in the RapidMiner application. Subsequently, the data underwent classification based on positive and negative sentiment scores. Following this, performance testing of the SVM algorithm with and without SMOTE was conducted to measure accuracy, precision, recall, f-measure, and AUC. This meticulous approach in the modeling phase reflects a rigorous and systematic effort to analyze sentiment patterns within the dataset, utilizing advanced algorithms to ascertain the effectiveness of SVM in sentiment classification, particularly in the presence of imbalanced data. The comparison of SVM performance with and without SMOTE provides valuable insights into the impact of addressing class imbalance on the algorithm's effectiveness. In conclusion, the modeling phase serves as a pivotal stage in evaluating the robustness of sentiment analysis techniques in discerning nuanced sentiment patterns related to climate change discourse within the tourism sector.

In the evaluation phase, the performance metrics, namely accuracy, precision, recall, f-measure, and AUC, of the SVM algorithm were systematically analyzed. The primary focus was on discerning the algorithm's effectiveness in accurately classifying sentiment within the dataset derived from public comments on National Geographic videos pertaining to climate change. The meticulous examination of these metrics provides a comprehensive understanding of the algorithm's proficiency in capturing both positive and negative sentiments. The AUC, in particular, serves as an essential indicator of the algorithm's discriminatory ability. This evaluative process contributes critical insights into the reliability and efficacy of the employed SVM algorithm, offering a robust foundation for interpreting the sentiment analysis results and advancing the understanding of public sentiment towards climate change within the tourism context.

In the deployment phase, a strategic analysis was undertaken to respond to the outcomes of the performance evaluation. The primary objective was to translate the insights derived from the sentiment analysis of public comments on National Geographic videos into actionable strategies. This involved delineating effective approaches for addressing the sentiments expressed by the audience regarding climate change within the tourism domain. The strategic deployment phase is pivotal in ensuring that the research findings are not only academically robust but also have practical implications. By aligning the sentiment analysis results with strategic responses, the deployment phase aims to contribute to informed decision-making processes and facilitate proactive measures within the tourism industry to address public concerns related to climate change. In conclusion, this phase serves as a bridge between research outcomes and practical applications, emphasizing the significance of translating academic insights into tangible strategies for real-world impact.

2.3 Support Vector Machine (SVM)

The SVM algorithm possesses distinct advantages and is highly relevant for sentiment classification based on datasets. Its ability to effectively handle high-dimensional data and nonlinear relationships makes it a preferred choice for sentiment analysis tasks. Additionally, SVM's robustness in handling binary and multiclass classification problems further enhances its applicability in diverse scenarios. The algorithm's versatility and proven performance in various research domains underscores its relevance and utility in sentiment analysis applications. In conclusion, leveraging SVM algorithms offers a powerful approach to accurately classify sentiment based on dataset characteristics, contributing to advancements in sentiment analysis methodologies. Meanwhile, the regression function of the SVM method is as follows.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

$$f(x) = w.x + b \quad (1)$$

Where $f(x)$ is the decision function, W is the weight vector perpendicular to the hyperplane, X is the input feature vector and B is the bias term. In the case of binary classification, the class label y_i of a data point x_i can be determined by the sign of $f(x)$:

$$y_i = \begin{cases} +1, \ if \ f(x_i) \ \geq 0 \\ -1, if \ f(x_i) \ < 1 \end{cases} \quad (2)$$

In non-linearly separable cases, SVM utilizes a kernel function $\mathrm{K}(x_i, x_j)$ to map the input feature vectors into a higher-dimensional space where the data becomes linearly separable. The decision function then becomes:

$$f(x) \ = \ \sum_{i=1}^{N} \propto_i, y_j \ \mathrm{K}(x_i, x) + \mathrm{b} \quad (3)$$

Where $\propto_i$ are the Lagrange multipliers obtained during training. The strengths and weaknesses of the SVM algorithm when employing the SMOTE operator provide valuable insights into its effectiveness in handling imbalanced datasets. One of SVM's notable advantages is its ability to effectively handle nonlinear classification tasks, making it suitable for sentiment analysis applications. In addition, SVM's margin maximization principle helps improve generalization performance, enhancing its robustness in handling complex datasets. However, when combined with the SMOTE operator, SVM may face computational complexity and scalability challenges, particularly with large-scale datasets. Furthermore, the effectiveness of SMOTE in addressing class imbalance relies heavily on the choice of parameters, which may affect the algorithm's performance and require careful optimization. In conclusion, while SVM with SMOTE offers benefits in addressing data imbalance, carefully considering its limitations and parameter tuning is essential to maximize its efficacy in sentiment analysis tasks.

3. RESULT AND DISCUSSION

Climate change is an issue of paramount importance that elicits both proponents and detractors [28]. Nevertheless, the impact of climate change on the sustainability of the tourism industry cannot be overstated [29]. This sector relies heavily on environmental stability and favorable weather conditions to thrive [30]. Consequently, any alteration in climate patterns significantly threatens its operation and profitability [31]. As such, addressing climate change is not merely an environmental concern but also a critical economic imperative for the tourism industry and the broader global economy [32].

The digital campaign on climate change has garnered a myriad of both support and opposition. In recent years, digital platforms have become instrumental in raising awareness about environmental issues, including climate change, reaching diverse audiences globally [33]. However, despite its potential to mobilize widespread action, digital campaigns often encounter criticism regarding their effectiveness in driving tangible change [34]. Critics argue that the virtual nature of these campaigns may dilute the sense of urgency or lead to "clicktivism" superficial engagement without meaningful impact [35]. Nevertheless, proponents emphasize the power of digital platforms to amplify voices, connect activists, and pressure policymakers, thereby catalyzing collective action on climate change [36]. In conclusion, while the digital campaign landscape offers vast opportunities for climate advocacy, its effectiveness ultimately hinges on strategic implementation and concerted efforts to translate online activism into real-world solutions.

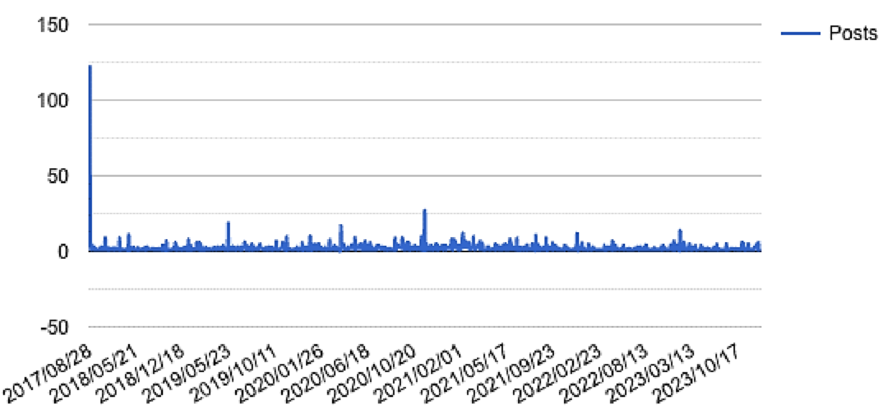

Figure 2. Number of Post Overtime

Based on public responses to content related to climate change, it is evident that public perceptions and reactions are highly diverse. The dissemination of information and discourse surrounding climate change often elicits a spectrum of reactions ranging from concern and urgency to skepticism and denial. While some segments

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

of the population strongly support climate action, others express skepticism toward the severity or even the existence of climate change. This variation in public perception underscores the complexity of addressing climate change and highlights the importance of nuanced communication strategies tailored to different audiences. In conclusion, understanding and navigating the diverse responses to climate change content is crucial for effective communication and fostering widespread engagement in climate mitigation and adaptation efforts.

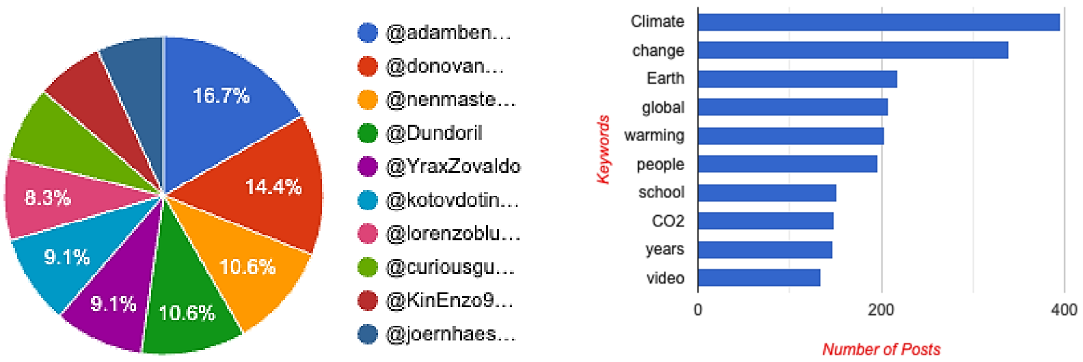

Figure 2. Most Frequently Used Words and Top Ten Posters

Based on the results of keyword identification from review data, the Netlytics website indicates that frequently occurring words include "climate," "change," "earth," "global," "warning," "people," "school," "CO2," "years," and "video." This comprehensive list reflects the recurring themes and topics prevalent in discussions surrounding climate change and its impacts. The prominence of "climate" and "change" underscores the central focus on environmental transformations. At the same time, words like "global" and "warning" suggest recognizing the widespread and urgent nature of climate-related challenges. Additionally, "school" and "video" hint at educational and multimedia strategies for communicating climate change issues. In conclusion, identifying these keywords provides valuable insights into the climate change discourse and can inform effective communication and engagement strategies with diverse audiences. Furthermore, the sentiment extract results on the Rapidminer application showed positive and negative sentiments regarding climate change video content posted by the National Geographic Channel.

Table 1. Extract Sentiment in Rapidminer

Review	Score	Total Score
stuff your globalist criminal lying frauds......chemtrails are artificially creating it....that and your criminal frauds of the plastics industry...you lying globalist propaganda frauds!!!!! Plant Marijuana.....STOP THESE FRAUDS AND THEIR LIES!	criminal (-0.62) lying (-0.62) frauds (-0.59) creating (0.31) criminal (-0.62) frauds (-0.59) lying (-0.62) propaganda (-0.26) frauds (-0.59) stop (-0.31) frauds (-0.59) lies (-0.46)	-5,53846153846154
I think the best solution is to give the people of the world something good to invest in that is secure and pays a great dividend. The developed nations of the world could provide backing and guarantees for an "Earth Bond" that pays 8% annual inflation-adjusted dividends with no limit to the guarantee and no ownership limit like was in place with war bonds. The bond funds would only be used for 2 purposes: 1) projects to improve the health of our planet. 2) provide employment, benefits and fair compensation to the world's poor Essentially you would employ the world's poor to improve the health of the planet and give plenty of incentive for everyone to invest in these endeavors through a guaranteed annual 8% inflation-adjusted dividend. Everybody wins and nobody feels like they were "taken from" like through taxation that never works.	best (0.82) solution (0.33) good (0.49) secure (0.36) great (0.79) backing (0.03) no (-0.31) guarantee (0.26) no (-0.31) like (0.38) war (-0.74) improve (0.49) benefits (0.41) fair (0.33) poor (-0.54) poor (-0.54) improve (0.49) incentive (0.38) wins (0.69) like (0.38) like (0.38)	4,58974358974359

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

Based on sentiment analysis, it is discernible that the most frequently occurring keywords in reviews of climate change content published by the National Geographic Channel complement those identified by the website Netlytic. The amalgamation of data from both sources provides a comprehensive understanding of the prevalent themes and sentiments surrounding climate change discourse. The keywords "climate" (393 occurrences), "change" (335 occurrences), "earth" (217 occurrences), "global" (209 occurrences), and "warming" (203 occurrences) featured prominently in both datasets, highlighting the overarching concern for environmental issues and the urgency to address climate change. Additionally, terms such as "people" (191 occurrences), "years" (146 occurrences), "school" (151 occurrences), "video" (133 occurrences), and "planet" (117 occurrences) indicate a focus on educating and engaging diverse audiences, mainly through multimedia platforms. The convergence of popular keywords underscores the interconnectedness of public sentiment and media representation in shaping perceptions and responses to climate change. In conclusion, leveraging insights from National Geographic Channel reviews and Netlytic data facilitates a holistic approach to understanding and addressing climate change's multifaceted challenges, as shown in the figure below.

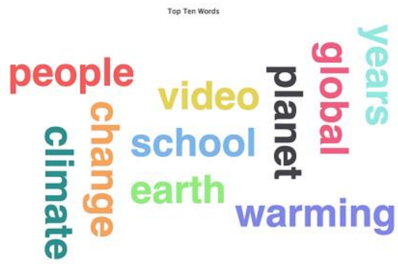

Figure 3. WordlCloud

In modeling using RapidMiner, connecting the SMOTE operator with the SVM algorithm allows for comparing performance outcomes. The utilization of the Synthetic Minority Over-sampling Technique (SMOTE) within the RapidMiner framework addresses the challenge of imbalanced datasets by generating synthetic samples of the minority class. By incorporating SMOTE with the Support Vector Machine (SVM) algorithm, which is renowned for its effectiveness in classification tasks, the modeling process aims to enhance predictive accuracy while mitigating the impact of class imbalance. This integration facilitates a robust model performance evaluation, enabling researchers and practitioners to make informed decisions regarding the most suitable approach for addressing imbalanced datasets in predictive modeling tasks.

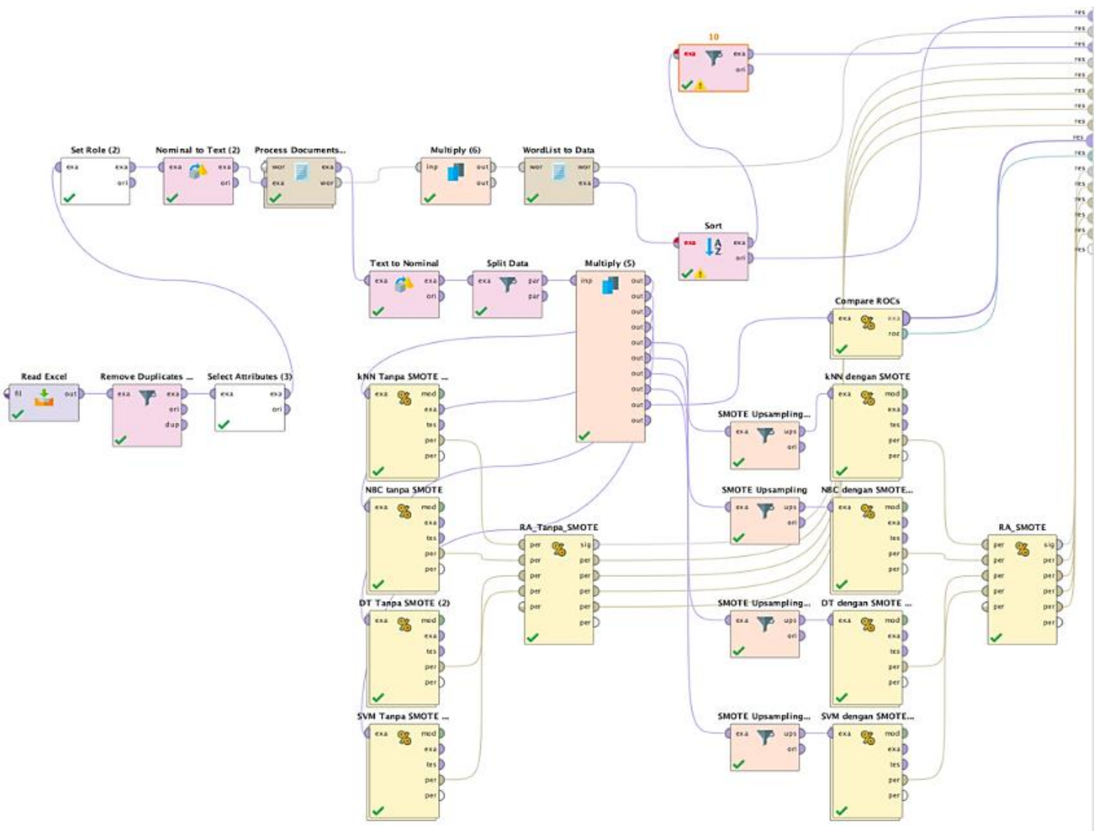

Figure 4. Modeling Process in Rapidminer

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

In the modeling process, the division between training and testing data follows a standard ratio of 70:30, allocating most of the dataset for training the model and a smaller portion for evaluating its performance. Additionally, to ensure data cleanliness and reduce redundancy, the data cleansing process involves removing duplicates, thereby enhancing the quality and reliability of the dataset. Furthermore, incorporating stopwords in the English language aids in refining the text data by filtering out commonly occurring words with minimal semantic value, thereby focusing the model's attention on more informative features. This meticulous approach to data preprocessing contributes to the robustness and effectiveness of the modeling process by enhancing the quality of input data and facilitating more accurate model predictions.

SVM Using SMOTE SVM Without using SMOTE

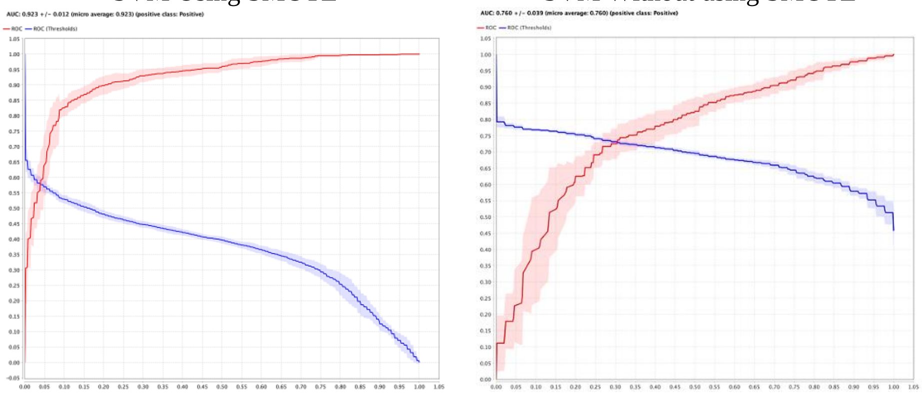

Figure 5. Area Under Curve (AUC) of the SVM Algorithm

Based on the values obtained from the confusion matrix, it is evident that SVM with SMOTE yields differing performance outcomes. Specifically, SVM with SMOTE demonstrates distinct performance metrics, indicating its effectiveness in handling imbalanced datasets. The Area Under the Curve (AUC) values, including optimistic, micro-average, and pessimistic scores, consistently demonstrate predictive solid performance, with micro-average AUC values ranging from 0.923 to 0.924. These results underscore the utility of integrating SMOTE with SVM for classification tasks, mainly when dealing with imbalanced datasets, as it enhances the model's ability to classify instances from minority classes accurately. In conclusion, the observed performance metrics validate the efficacy of SVM with SMOTE in improving classification performance, thus emphasizing its relevance in data-driven decision-making processes.

Table 2. SVM Performance in Sentiment Classification

SVM Using SMOTE	SVM Without using SMOTE
PerformanceVector: accuracy: 86.15% +/- 1.68% (micro average: 86.15%) ConfusionMatrix: True: Negative Positive Negative: 1082 158 Positive: 196 1120 AUC (optimistic): 0.924 +/- 0.012 (micro average: 0.924) (positive class: Positive) AUC: 0.923 +/- 0.012 (micro average: 0.923) (positive class: Positive) AUC (pessimistic): 0.923 +/- 0.012 (micro average: 0.923) (positive class: Positive) precision: 85.17% +/- 2.16% (micro average: 85.11%) (positive class: Positive) ConfusionMatrix: True: Negative Positive Negative: 1082 158 Positive: 196 1120 recall: 87.64% +/- 3.39% (micro average: 87.64%) (positive class: Positive) ConfusionMatrix: True: Negative Positive Negative: 1082 158 Positive: 196 1120	PerformanceVector: accuracy: 74.23% +/- 0.55% (micro average: 74.23%) ConfusionMatrix: True: Negative Positive Negative: 10 3 Positive: 443 1275 AUC (optimistic): 0.764 +/- 0.038 (micro average: 0.764) (positive class: Positive) AUC: 0.760 +/- 0.039 (micro average: 0.760) (positive class: Positive) AUC (pessimistic): 0.757 +/- 0.040 (micro average: 0.757) (positive class: Positive) precision: 74.22% +/- 0.50% (micro average: 74.21%) (positive class: Positive) ConfusionMatrix: True: Negative Positive Negative: 10 3 Positive: 443 1275 recall: 99.77% +/- 0.38% (micro average: 99.77%) (positive class: Positive) ConfusionMatrix: True: Negative Positive Negative: 10 3 Positive: 443 1275

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

SVM Using SMOTE	SVM Without using SMOTE
f_measure: 86.34% +/- 1.79% (micro average: 86.35%) (positive class: Positive) ConfusionMatrix: True: Negative Positive Negative: 1082 158 Positive: 196 1120	f_measure: 85.11% +/- 0.30% (micro average: 85.11%) (positive class: Positive) ConfusionMatrix: True: Negative Positive Negative: 10 3 Positive: 443 1275

Based on the values obtained from the confusion matrix, it is evident that SVM without SMOTE exhibits specific performance metrics indicative of its classification accuracy and effectiveness. The accuracy rate, averaging 74.23% with a micro average of 74.23%, suggests a moderate level of overall correct classifications. However, further examination of precision, recall, and F-measure metrics reveals a nuanced understanding of the model's performance. While precision demonstrates the proportion of correctly classified positive instances among all instances predicted as positive, recall represents the proportion of correctly classified positive instances among all actual positive instances. The F-measure, which harmonizes precision and recall, provides a balanced assessment of the model's performance. In this instance, the precision (74.22%), recall (99.77%), and F-measure (85.11%) values reflect reasonably high levels of correct classifications for the positive class, indicating the model's proficiency in identifying positive instances. Thus, despite not employing SMOTE to address the class imbalance, SVM demonstrates a commendable ability to classify positive instances accurately.

Based on the values derived from the confusion matrix, it is evident that SVM utilizing SMOTE demonstrates notable performance metrics, particularly classification accuracy. The model exhibits a high overall correct classification with an accuracy rate averaging 86.15% and a micro average of 86.15%. Further precision, recall, and F-measure metrics analysis reveal additional model performance insights. Precision, representing the proportion of correctly classified positive instances among all instances predicted as positive, yields an average of 85.17%. Additionally, recall, indicating the proportion of correctly classified positive instances among all actual positive instances, averages 87.64%. The F-measure, synthesizing precision and recall, provides a balanced assessment of the model's performance, averaging 86.34%. These metrics collectively underscore the effectiveness of SVM with SMOTE in accurately classifying positive instances, thereby affirming its utility in addressing class imbalance and enhancing predictive performance.

Comparing the results of SVM with and without SMOTE reveals distinct performance disparities. When utilizing SMOTE, SVM demonstrates superior classification accuracy, averaging 86.15% with a micro average of 86.15%, compared to 74.23% accuracy without SMOTE. Precision, recall, and F-measure metrics also exhibit significant improvements with SMOTE, averaging 85.17%, 87.64%, and 86.34%, respectively, compared to 74.22%, 99.77%, and 85.11% without SMOTE. Additionally, the AUC values for SVM with SMOTE remain consistently high, with optimistic, micro average, and pessimistic scores ranging from 0.923 to 0.924, compared to slightly lower AUC values of 0.757 to 0.764 without SMOTE. These findings underscore the efficacy of SMOTE in addressing class imbalance and enhancing the predictive performance of SVM, particularly in accurately classifying positive instances.

The analysis of the results reveals a substantial disparity in the performance of SVM with and without SMOTE. SVM with SMOTE demonstrates a notable increase in classification accuracy, achieving an average of 86.15% with a micro average of 86.15%, compared to 74.23% accuracy without SMOTE. Additionally, precision improves from 74.22% without SMOTE to 85.17% with SMOTE, recall increases from 99.77% to 87.64%, and the F-measure rises from 85.11% to 86.34%. Furthermore, the Area Under the Curve (AUC) values consistently remain high for SVM with SMOTE, ranging from 0.923 to 0.924, compared to slightly lower AUC values of 0.757 to 0.764 without SMOTE. These findings underscore the efficacy of SMOTE in enhancing SVM's performance, particularly in handling imbalanced datasets, and emphasize the significance of employing appropriate techniques for optimal predictive modeling outcomes.

4. CONCLUSION

The research findings present a comparative analysis of the SVM algorithm's performance with and without SMOTE. The results indicate that SVM without SMOTE achieved an accuracy of 74.23% +/- 0.55%, precision of 74.22% +/- 0.50%, recall of 99.77% +/- 0.38%, f_measure of 85.11% +/- 0.30%, and AUC (optimistic) of 0.764 +/- 0.038. Conversely, SVM with SMOTE obtained an accuracy of 86.15% +/- 1.68%, precision of 85.17% +/- 2.16%, recall of 87.64% +/- 3.39%, f_measure of 86.34% +/- 1.79%, and AUC of 0.923 +/- 0.012. These results highlight the notable performance improvements when employing SMOTE, demonstrating enhanced accuracy, precision, recall, f_measure, and AUC. Using SMOTE contributes to mitigating the class imbalance issue and consequently improves the SVM algorithm's effectiveness in sentiment classification tasks related to climate change and tourism. Thus, integrating SMOTE into the SVM algorithm is promising for achieving more robust sentiment analysis outcomes in such contexts. The practical implications of this research are multifaceted and extend beyond academia into real-world applications. The findings highlight the importance of addressing class imbalance in predictive modeling tasks, particularly in domains where minority classes are

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

prevalent. By demonstrating the efficacy of the SMOTE technique in enhancing the performance of Support Vector Machine (SVM) models, this research provides practitioners with a valuable tool for improving the accuracy and reliability of classification systems. Moreover, the observed improvements in classification metrics such as accuracy, precision, recall, and F-measure underscore the potential for SMOTE-enhanced SVM models to be deployed in various practical settings, including healthcare, finance, and cybersecurity, where accurate classification of minority instances is crucial for decision-making processes. In conclusion, this research contributes to advancing the field of machine learning and offers actionable insights that can positively impact decision-making and problem-solving in real-world scenarios.

REFERENCES

[1] D. Navarro-Drazich *et al.*, "Climate change and tourism in South and Central America," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–17, 2023, doi: 10.1080/09669582.2023.2210783.
[2] G. D. Sharma, B. Taheri, R. Chopra, and J. S. Parihar, "Relationship between climate change and tourism: an integrative review," *Serv. Ind. J.*, vol. 0, no. 0, pp. 1–28, 2023, doi: 10.1080/02642069.2023.2254714.
[3] K. Dube *et al.*, "Tourism and climate change in Africa: informing sector responses," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–21, 2023, doi: 10.1080/09669582.2023.2193355.
[4] F. Wolf *et al.*, "Small island developing states, tourism and climate change," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–19, 2022, doi: 10.1080/09669582.2022.2112203.
[5] C. T. Cavaliere and L. J. Ingram, "Climate change and anger: misogyny and the dominant growth paradigm in tourism," *Ann. Leis. Res.*, vol. 26, no. 3, pp. 354–371, 2023, doi: 10.1080/11745398.2021.1949732.
[6] S. Becken, E. Whittlesea, J. Loehr, and D. Scott, "Tourism and climate change: evaluating the extent of policy integration," *J. Sustain. Tour.*, vol. 28, no. 10, pp. 1603–1624, 2020, doi: 10.1080/09669582.2020.1745217.
[7] M. Rutty, M. Hewer, N. Knowles, and S. Ma, "Tourism & climate change in North America: regional state of knowledge," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–24, 2022, doi: 10.1080/09669582.2022.2127742.
[8] A. Arabadzhyan, P. Figini, C. García, M. M. González, Y. E. Lam-González, and C. J. León, "Climate change, coastal tourism, and impact chains–a literature review," *Curr. Issues Tour.*, vol. 24, no. 16, pp. 2233–2268, 2021, doi: 10.1080/13683500.2020.1825351.
[9] M. Alizadeh, R. Mirzaei, and A. Dittmann, "Climate change and its potential impacts on sustainable tourism development," *Anatolia*, vol. 32, no. 3, pp. 443–455, 2021, doi: 10.1080/13032917.2021.1886130.
[10] R. Steiger, N. Knowles, K. Pöll, and M. Rutty, "Impacts of climate change on mountain tourism: a review," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–34, 2022, doi: 10.1080/09669582.2022.2112204.
[11] C. Michael Hall and J. Saarinen, "20 years of Nordic climate change crisis and tourism research: a review and future research agenda," *Scand. J. Hosp. Tour.*, vol. 21, no. 1, pp. 102–110, 2021, doi: 10.1080/15022250.2020.1823248.
[12] H. Zhao and A. Ewert, "College Students' Knowledge and Perceptions of Tourism Climate Change Impacts: Do Major, Grade and Gender Matter?," *J. Hosp. Tour. Educ.*, vol. 33, no. 4, pp. 258–269, 2021, doi: 10.1080/10963758.2020.1727342.
[13] D. Scott and S. Gössling, "From Djerba to Glasgow: have declarations on tourism and climate change brought us any closer to meaningful climate action?," *J. Sustain. Tour.*, vol. 30, no. 1, pp. 199–222, 2021, doi: 10.1080/09669582.2021.2009488.
[14] L. Horne, S. De Urioste-Stone, P. Rahimzadeh Bajgiran, and E. Seekamp, "Understanding Tourism Suppliers' Resilience to Climate Change in a Rural Destination in Maine," *Tour. Plan. Dev.*, vol. 0, no. 0, pp. 1–22, 2022, doi: 10.1080/21568316.2022.2083222.
[15] R. Steiger, E. Posch, G. Tappeiner, and J. Walde, "Seasonality matters: simulating the impacts of climate change on winter tourism demand," *Curr. Issues Tour.*, vol. 26, no. 17, pp. 2777–2793, 2023, doi: 10.1080/13683500.2022.2097861.
[16] W. Mushawemhuka, J. M. Fitchett, and G. Hoogendoorn, "Climate change and adaptation in the Zimbabwean nature-based tourism industry," *Anatolia*, vol. 00, no. 00, pp. 1–12, 2022, doi: 10.1080/13032917.2022.2132412.
[17] A. McCreary, E. Seekamp, L. R. Larson, J. Smith, and M. A. Davenport, "Climate Change and Nature-Based Tourism: How Do Different Types of Visitors Respond?," *Tour. Plan. Dev.*, 2020, doi: 10.1080/21568316.2020.1861079.
[18] S. Wang, Y. Yu, J. Chen, and J. Liu, "Impact of climate change on cherry blossom viewing tourism: analysis and simulation based on Weibo proxy data," *Curr. Issues Tour.*, vol. 26, no. 5, pp. 718–734, 2023, doi: 10.1080/13683500.2022.2049711.
[19] Y. A. Singgalen, "Analisis Sentimen Top 10 Traveler Ranked Hotel di Kota Makassar Menggunakan Algoritma Decision Tree dan Support Vector Machine," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 323–332, 2023, doi: 10.30865/klik.v4i1.1153.
[20] Y. A. Singgalen, "Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1870–

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 2, February 2024, Page 357-367
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i2.4908

1882, 2023, doi: 10.47065/bits.v4i4.3199.
[21] Y. A. Singgalen, "Analisis Sentimen Konsumen terhadap Food , Services , and Value di Restoran dan Rumah Makan Populer Kota Makassar Berdasarkan Rekomendasi Tripadvisor Menggunakan Metode CRISP-DM dan," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1899–1914, 2023, doi: 10.47065/bits.v4i4.3231.
[22] Y. A. Singgalen, "Analisis Sentimen Wisatawan terhadap Taman Nasional Bunaken dan Top 10 Hotel Rekomendasi Tripadvisor Menggunakan Algoritma SVM dan DT berbasis CRISP-DM," *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 367–379, 2023, doi: 10.47065/josyc.v4i2.3092.
[23] Y. A. Singgalen, "Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree (DT)," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, pp. 237–248, 2023, doi: 10.30865/json.v5i2.7081.
[24] Y. A. Singgalen, "Analisis Performa Algoritma NBC , DT , SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1634–1646, 2022, doi: 10.47065/bits.v4i3.2766.
[25] Y. A. Singgalen, "Analisis Perilaku Wisatawan Berdasarkan Data Ulasan di Website Tripadvisor Menggunakan CRISP-DM : Wisata Minat Khusus Pendakian Gunung Rinjani dan Gunung Bromo," *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 326–338, 2023, doi: 10.47065/josyc.v4i2.3042.
[26] Y. A. Singgalen, "Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM," *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 614–625, 2023, doi: 10.47065/josh.v4i2.2999.
[27] R. Obiedat *et al.*, "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.
[28] D. Scott, N. L. B. Knowles, S. Ma, M. Rutty, and R. Steiger, "Climate change and the future of the Olympic Winter Games: athlete and coach perspectives," *Curr. Issues Tour.*, vol. 26, no. 3, pp. 480–495, 2023, doi: 10.1080/13683500.2021.2023480.
[29] H. Rice, S. Cohen, D. Scott, and R. Steiger, "Climate change risk in the Swedish ski industry," *Curr. Issues Tour.*, vol. 25, no. 17, pp. 2805–2820, 2022, doi: 10.1080/13683500.2021.1995338.
[30] A. Rai, D. P. Ayadi, B. Shrestha, and A. Mishra, "On the realities of gender inclusion in climate change policies in Nepal," *Policy Des. Pract.*, vol. 4, no. 4, pp. 501–516, 2021, doi: 10.1080/25741292.2021.1935643.
[31] C. J. León, E. Giannakis, G. Zittis, D. Serghides, Y. E. Lam-González, and C. García, "Tourists' Preferences for Adaptation Measures to Build Climate Resilience at Coastal Destinations. Evidence from Cyprus," *Tour. Plan. Dev.*, vol. 20, no. 6, pp. 973–999, 2023, doi: 10.1080/21568316.2021.1958914.
[32] Y. Fang, A. Trupp, J. S. Hess, and S. Ma, "Tourism under climate crisis in Asia: impacts and implications," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–17, 2022, doi: 10.1080/09669582.2022.2112202.
[33] M. Sheller, "Reconstructing tourism in the Caribbean: connecting pandemic recovery, climate resilience and sustainable tourism through mobility justice," *J. Sustain. Tour.*, vol. 29, no. 9, pp. 1–14, 2020, doi: 10.1080/09669582.2020.1791141.
[34] H. Olya, N. Kim, and M. J. Kim, "Climate change and pro-sustainable behaviors: application of nudge theory," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–19, 2023, doi: 10.1080/09669582.2023.2201409.
[35] D. Scott, N. Knowles, and R. Steiger, "Is snowmaking climate change maladaptation?," *J. Sustain. Tour.*, vol. 32, no. 2, pp. 282–303, 2022, doi: 10.1080/09669582.2022.2137729.
[36] D. Zheng, C. Huang, and B. Oraltay, "Digital cultural tourism: progress and a proposed framework for future research," *Asia Pacific J. Tour. Res.*, vol. 28, no. 3, pp. 234–253, 2023, doi: 10.1080/10941665.2023.2217958.