Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

Sentiment Classification of S.E.A Aquarium Singapore Reviews through CRISP-DM using DT and SVM with SMOTE

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
Email Penulis Korespondensi: yerik.afrianto@atmajaya.ac.id
Submitted: **21/12/2023**; Accepted: **30/12/2023**; Published**: 30/12/2023**

Abstract−In recent years, sentiment analysis has emerged as a critical area of research due to its wide-ranging applications in understanding public opinion, customer feedback, and social media sentiment. However, one of the significant challenges faced in sentiment analysis is the handling of imbalanced datasets, where the distribution of sentiment classes is uneven, leading to biased model performance. This study employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to investigate sentiment analysis algorithms, mainly focusing on the Support Vector Machine (SVM) algorithm and the integration of the Synthetic Minority Over-sampling Technique (SMOTE). Through systematic experimentation and evaluation, the research demonstrates the superior performance of the SVM-SMOTE model in handling imbalanced datasets, achieving an accuracy of 98.46%, an AUC of 1.000, precision of 100.00%, recall of 96.91%, and an impressive F-measure of 98.42%. Additionally, the evaluation unveils specific toxicity scores across various categories, with Toxicity scoring at 0.11036 and 0.93915, Severe Toxicity at 0.00905 and 0.45895, Identity Attack at 0.02415 and 0.66373, Insult at 0.05149 and 0.85793, Profanity at 0.06392 and 0.93426, and Threat at 0.01562 and 0.51957. These numerical indicators provide quantitative insights into potential harm within analyzed content, emphasizing the efficacy of the SVM-SMOTE model in real-world applications and contributing to the advancement of sentiment analysis within the CRISP-DM framework.

Keywords: CRISP-DM; DT; Sentiment Classification; SMOTE; SVM

1. INTRODUCTION

Artificial tourist destinations have gained significant popularity, prompting thorough analysis from travelers' perspectives to identify and analyze preferences and behaviors [1], [2], [3]. The emergence of such destinations as key attractions underscores the need for meticulous examination to comprehend the underlying factors driving the allure and appeal [4], [5], [6]. Through meticulous scrutiny of tourist behaviors and preferences, insights into the dynamics shaping the tourism landscape are gleaned, facilitating informed decision-making and strategic planning within the industry [7]. Consequently, the analysis of artificial tourist destinations serves as an indispensable tool for stakeholders aiming to capitalize on emerging trends and meet the evolving demands of contemporary travelers.

One of Singapore's popular artificial tourist destinations is the S.E.A Aquarium on Sentosa Island. Entering and exploring the marine realm of the S.E.A. Aquarium allows visitors to immerse themselves in a captivating experience, encountering over 100,000 marine animals comprising more than 1,000 species across 50 different habitats. This vast array of marine life offers an unparalleled opportunity for education and exploration, making it an unforgettable experience for visitors. The diversity of species and habitats within the aquarium underscores its significance as a prominent attraction, contributing to Singapore's vibrant tourism landscape and serving as a testament to the marvels of marine biodiversity.

This study aims to analyze the sentiment of tourists through data sourced from reviews on TripAdvisor and influencer channels on YouTube. This endeavor is essential for comprehending the behavioral context of tourists visiting the S.E.A Aquarium in Singapore. A comprehensive understanding of tourists' sentiments towards the attraction is attained by examining reviews from multiple platforms, including user-generated content and influencer perspectives [8], [9], [10], [11]. This analysis enables stakeholders to gain insights into the factors influencing tourists' perceptions and experiences at the S.E.A Aquarium, informing strategic decisions to enhance visitor satisfaction and promote sustainable tourism development.

The urgency of this research lies in its potential to offer invaluable insights into the dynamics of tourist sentiment towards the S.E.A Aquarium in Singapore. By analyzing data derived from TripAdvisor reviews and influencer channels on YouTube, this study seeks to unravel the underlying factors shaping tourists' perceptions and experiences in the attraction [12], [13], [14]. Such insights are crucial for stakeholders in the tourism industry, as they enable informed decision-making to optimize visitor satisfaction and foster sustainable tourism practices [15], [16], [17], [18], [19], [20], [21]. Consequently, this research promises to drive positive outcomes for both the S.E.A Aquarium and the broader tourism ecosystem in Singapore.

This research's theoretical and practical implications are profound, extending beyond the immediate scope of analyzing tourist sentiment towards the S.E.A Aquarium in Singapore. By delving into the nuanced dynamics of visitor perceptions and experiences, this study contributes to the enrichment of existing theoretical frameworks in tourism studies, particularly in understanding the interplay between user-generated content and influencer narratives in shaping tourist behavior [22], [23], [24], [25], [26]. Moreover, the findings of this research hold practical

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

significance for stakeholders in the tourism industry, offering actionable insights to inform strategic decision-making processes [27], [28], [29], [30]. By leveraging the insights from analyzing TripAdvisor reviews and influencer content on YouTube, stakeholders devise targeted marketing strategies, enhance visitor experiences, and optimize resource allocation to maximize the attractiveness and sustainability of the S.E.A Aquarium as a key tourist destination. This research bridges the gap between theory and practice, offering tangible benefits for academic discourse and industry practitioners.

Exploring similar research endeavors presents a fertile ground for expanding knowledge and fostering collaborative learning within tourism studies. This research identifies gaps, trends, and emerging themes that warrant further exploration by examining related studies investigating tourist sentiment toward various attractions or destinations [31], [32]. This process enhances the breadth and depth of existing knowledge and offers an opportunity to contribute novel insights and methodologies to the academic discourse. Through interdisciplinary collaboration and knowledge exchange, this research collectively advances our understanding of tourist behavior, preferences, and experiences, enriching theoretical frameworks and informing practical applications in the tourism industry [33], [34], [35]. Consequently, engaging with similar research broadens the intellectual horizon and fosters a culture of innovation and continuous improvement in addressing complex challenges within tourism studies.

The limitation of this research primarily hinges on the constraints posed by the methodology employed, the availability and adequacy of data resources, and the contextual specificity of the S.E.A Aquarium in Singapore. While the CRISP-DM framework provides a systematic approach to data analysis, its efficacy is contingent upon the quality and comprehensiveness of the data sources utilized. Furthermore, the inherent limitations of relying on reviews from platforms like TripAdvisor and influencer channels on YouTube may introduce biases or overlook certain aspects of tourist sentiment. Additionally, the unique contextual factors surrounding the S.E.A Aquarium, such as cultural nuances, visitor demographics, and seasonal variations, may not be fully captured within the scope of this study. Despite these limitations, acknowledging and addressing these methodological and contextual constraints is essential for ensuring the robustness and validity of the research findings, thereby facilitating more nuanced insights into tourist behavior and experiences at the S.E.A Aquarium in Singapore.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis

Conducting a gap analysis is imperative to pinpoint areas of novelty and potential contributions, both theoretically and practically, within the research domain. This research identifies gaps, inconsistencies, or unexplored avenues that warrant further investigation by systematically assessing existing literature and empirical studies. This analytical process illuminates areas where current knowledge falls short and unveils opportunities for theoretical advancement and practical application. By meticulously examining the research landscape, scholars strategically position the work to address critical gaps and make meaningful contributions to advancing knowledge in respective fields. In essence, the pursuit of gap analysis catalyzes fostering innovation, driving intellectual progress, and shaping the future trajectory of research endeavors.

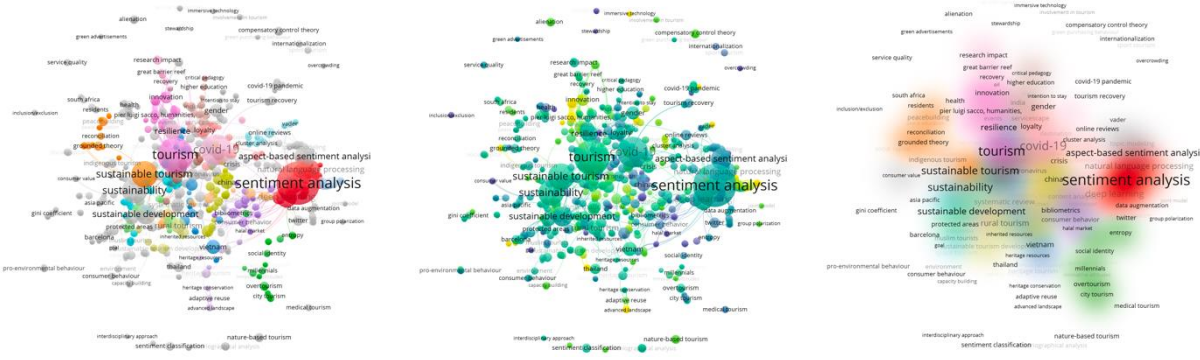

Figure 1. Gap Analysis using VosViewer

Figure 1 shows the overlay, density, and network visualization of topics related to man-made tourism. Based on the theoretical gap analysis, the study of sentiment analysis has gained prominence; however, there remains a paucity of research focusing specifically on tourist sentiment at tourist destinations. While sentiment analysis has garnered attention across various domains for its ability to extract insights from textual data, its application within the context of tourism remains relatively underexplored. Despite the burgeoning interest in understanding tourist behaviors and preferences, there is a notable gap in the literature concerning the sentiment tourists express towards specific attractions or destinations. This gap presents an opportunity to delve into the nuances of tourist sentiment and its implications for destination management, marketing strategies, and visitor experience enhancement. In essence,

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

bridging this theoretical gap holds promise for enriching our understanding of tourist behavior and contributing to the advancement of sentiment analysis within the context of tourism studies.

Several studies have highlighted sustainability as a prominent topic within tourism. The relevance of this topic extends to sentiment classification research, mainly through aspect-based sentiment analysis, which centers on assessing sentiment toward specific tourism activities. Sustainability has emerged as a critical concern in tourism discourse, with stakeholders increasingly emphasizing responsible practices to mitigate environmental impact and preserve cultural integrity. Aspect-based sentiment analysis offers a nuanced approach to understanding tourists' sentiments towards sustainability initiatives and related activities, shedding light on perceptions, preferences, and behaviors. By integrating sustainability considerations into sentiment classification frameworks, this research provides valuable insights that inform sustainable tourism development strategies and contribute to the long-term viability of tourism destinations. The intersection of sustainability and sentiment analysis presents a fertile ground for advancing knowledge and fostering sustainable practices within the tourism industry.

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

The CRISP-DM framework is adopted as a solution-oriented approach to address the research problem in sentiment classification using Support Vector Machine (SVM) and Decision Tree (DT) models. CRISP-DM provides a structured methodology comprising distinct phases such as business understanding, data understanding, data preparation, modeling, evaluation, and deployment. By leveraging CRISP-DM, this research systematically navigates the complexities of sentiment classification, from understanding the business context to deploying predictive models effectively. Adopting CRISP-DM ensures methodological rigor and consistency, facilitating the efficient development and implementation of SVM and DT models for sentiment analysis. Consequently, this framework is a robust guide in tackling the research problem, enabling them to generate reliable and actionable insights from textual data for informed decision-making and strategic planning.

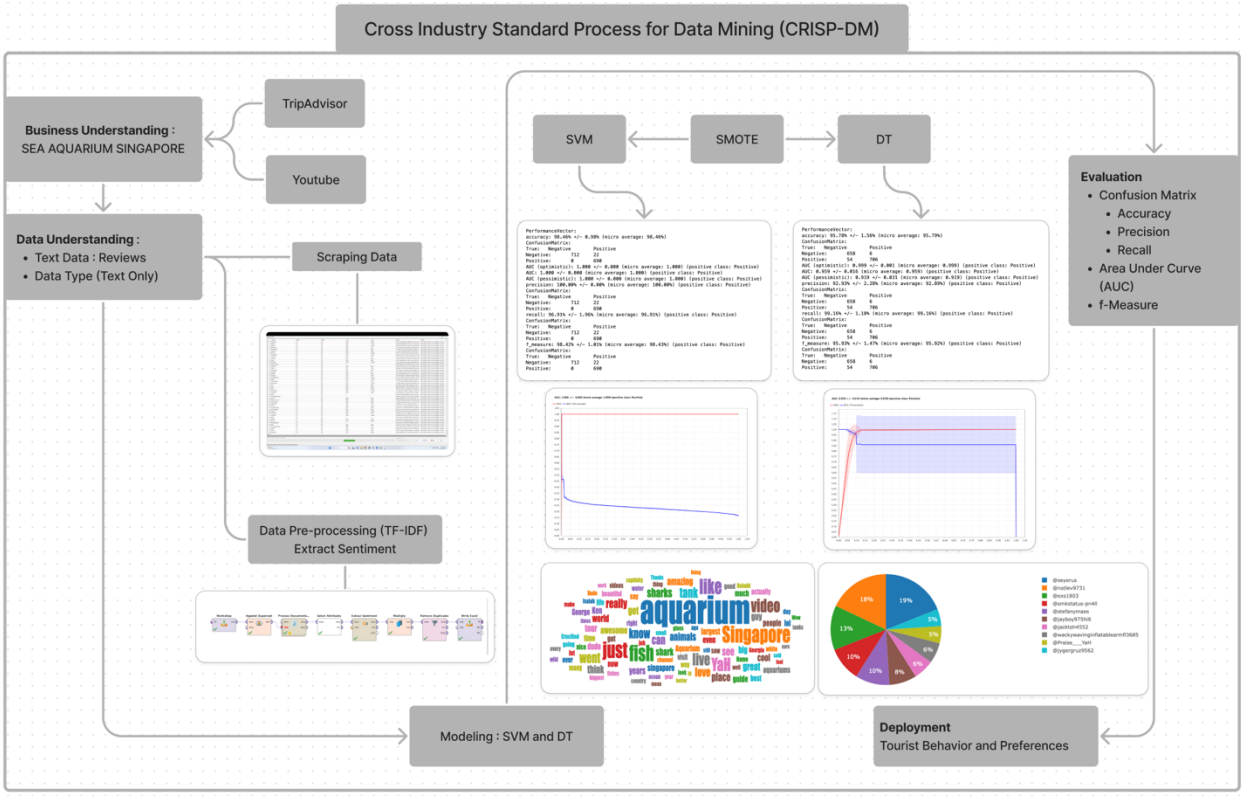

Figure 2. Cross-Industry Standard Process for Data Mining (CRISP-DM) Framework

Figure 2 shows the implementation of the CRISP-DM framework. Based on the processed textual data from TripAdvisor reviews, 1071 data entries are available for analysis, indicating a substantial volume of user-generated content to draw insights. Concurrently, the dataset obtained from YouTube, identified by the unique video identifier "nzMSLssodPY," comprises 2001 textual entries, further enriching the corpus for sentiment analysis. This abundance of data from diverse sources presents a valuable opportunity to conduct comprehensive and robust analyses of tourist sentiment toward the designated attraction. The sizable datasets from TripAdvisor and YouTube offer a comprehensive perspective on visitor experiences and perceptions, laying the groundwork for insightful findings and informed decision-making in tourism studies.

Using CRISP-DM, the sentiment analysis results obtained using Communalytic and RapidMiner will be compared and analyzed to generate pertinent recommendations. CRISP-DM provides a structured data mining and

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

analysis framework, encompassing distinct phases such as business understanding, data understanding, data preparation, modeling, evaluation, and deployment. By adopting this systematic approach, this research ensures methodological rigor and consistency in comparing and analyzing sentiment analysis outcomes from different tools or methodologies. Using Communalytic and RapidMiner for sentiment analysis enables this research to leverage diverse analytical techniques and algorithms, thereby facilitating a comprehensive assessment of tourist sentiment toward the designated attraction. Consequently, the comparative analysis facilitated by CRISP-DM holds promise for generating actionable insights and recommendations that contribute to informed decision-making and strategic planning in the tourism industry.

2.2.1 Business Understanding

In the business understanding phase, it is imperative to delineate the specific context of inquiry, emphasizing man-made tourism, mainly focusing on the S.E.A Aquarium in Singapore, which will be analyzed based on reviews from TripAdvisor and YouTube platforms. This initial phase serves as a crucial foundation for the subsequent stages of the CRISP-DM framework, elucidating the key objectives, stakeholders, and contextual nuances pertinent to the research endeavor. By precisely delineating the scope and context of the study, ensure alignment with the overarching research goals and maximize the relevance and applicability of the subsequent analyses and recommendations. The business understanding phase lays the groundwork for a comprehensive and systematic exploration of tourist sentiment towards the designated attraction, facilitating informed decision-making and strategic planning in the tourism industry.

The selection of data sources from the YouTube and TripAdvisor platforms considers the substantial volume of textual data, exceeding 1000 entries. Consequently, this abundance of data provides a robust foundation for interpreting the perceptions and preferences of tourists visiting the designated destination. By leveraging data from platforms with extensive user-generated content, such as YouTube and TripAdvisor, this research captures diverse opinions and experiences, offering valuable insights into the multifaceted nature of tourist sentiment. This strategic approach to data selection enhances the depth and breadth of the analysis, enabling a comprehensive understanding of the factors shaping tourist perceptions and behaviors at the designated attraction.

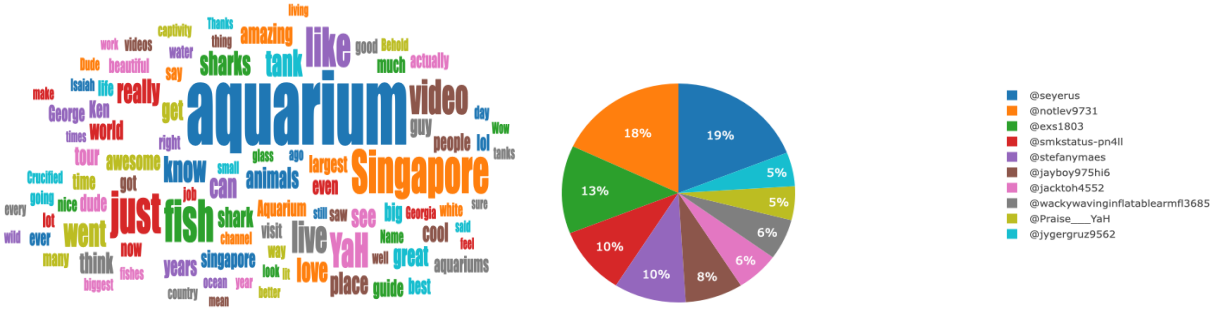

Figure 3. Frequently Used Words of Video Reviews in Communalytic

Figure 3 shows the word clouds in communistic. Based on the identification of frequently used words from 2001 comments on the video with the ID "nzMSLssodPY," it is evident that specific terms prominently feature in the discourse surrounding the S.E.A Aquarium in Singapore. Words such as "aquarium" (253 occurrences), "Singapore" (138 occurrences), "fish" (137 occurrences), and "sharks" (67 occurrences) emerge as recurrent themes, reflecting the central focus on marine life and the destination itself. Additionally, terms like "just" (134 occurrences), "like" (124 occurrences), and "really" (66 occurrences) suggest expressions of personal opinion or emphasis, indicating the subjective nature of the comments. The prevalence of words such as "love" (67 occurrences), "great" (55 occurrences), and "place" (61 occurrences) further underscores positive sentiments towards the aquarium experience. Conversely, terms like "can" (66 occurrences) and "know" (73 occurrences) may indicate inquiries or expressions of curiosity from viewers. Overall, the frequency analysis offers valuable insights into the predominant themes and sentiments expressed within the comments, providing a basis for further qualitative analysis and interpretation of tourist perceptions and experiences at the S.E.A Aquarium in Singapore.

Further analysis of the top-ten posters reveals vital contributors to the discourse surrounding the S.E.A Aquarium in Singapore. The usernames @seyerus (19 posts), @notlev9731 (18 posts), and @exs1803 (13 posts) emerge as the most prolific contributors, indicating significant engagement with the topic. Additionally, usernames such as @smkstatus-pn4ll (10 posts) and @stefanymaes (10 posts) demonstrate notable discussion participation. While some users, such as @jayboy975hi6 (8 posts) and @jacktoh4552 (6 posts), contribute moderately, others like @wackywaviginflatablearmfl3685 (6 posts), @Praise___YaH (5 posts), and @jygergruz9562 (5 posts) contribute to a lesser extent. This analysis sheds light on the distribution of contributions within the online discourse, highlighting key influencers and respective levels of engagement.

Subsequently, an observational study is conducted at the S.E.A Aquarium to reaffirm the patterns of interaction and behavior exhibited by tourists based on various categories, including solo travelers, couples, business visitors, families, and friends. This observational approach offers valuable insights into visitor engagement and social dynamics within the attraction, allowing for a deeper understanding of distinct visitor groups' differing needs and

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

preferences. By observing the behaviors and interactions of tourists in different contexts, this research augments analysis with qualitative observations, providing a holistic perspective on the visitor experience at the S.E.A Aquarium.

Figure 4. Field Observation in S.E.A Aquarium Singapore

Figure 4 shows the field observation of the S.E.A Aquarium in Singapore. Based on the observational findings, visitors to the S.E.A Aquarium are classified into distinct categories: solo travelers, couples, friends, families, and business visitors. Consequently, there is a need to collect and categorize review data from TripAdvisor alongside feedback from YouTube, specifically through influencer George Mavrakis' video with the ID "nzMSLssodPY," which has garnered a significant viewership of 3,950,962 views as of May 20, 2020. This comprehensive approach to data collection ensures the inclusion of diverse perspectives and experiences, allowing for a nuanced analysis of tourist sentiment and preferences towards the destination across different visitor types.

Based on the identification of the number of reviews for the S.E.A Aquarium on TripAdvisor, a total of 6,852 reviews were identified, categorized as follows: Excellent (3,664 reviews), Excellent (2,235 reviews), Average (726 reviews), Poor (151 reviews), and Terrible (76 reviews). This comprehensive dataset provides a quantitative overview of visitor feedback, allowing for a detailed analysis of the overall satisfaction levels and sentiments expressed towards the attraction. The distribution of reviews across different rating categories offers insights into the strengths and weaknesses of the S.E.A Aquarium, informing strategic decision-making and improvement initiatives to enhance the visitor experience.

Therefore, the text data downloaded from TripAdvisor and YouTube was preprocessed and extracted to be classified based on negative and positive sentiments. This preprocessing step involves cleaning the data to remove noise and irrelevant information, such as punctuation marks and stop words while standardizing the text format for consistency in analysis. Subsequently, sentiment analysis techniques, such as lexicon-based or machine-learning approaches, were applied to classify each review's sentiment polarity as negative or positive. By categorizing the reviews based on sentiment, this research gains valuable insights into tourists' overall perception and satisfaction levels towards the SEA Aquarium, facilitating informed decision-making and strategic planning to address areas of concern and capitalize on strengths.

2.2.2 Data Understanding

During the data understanding phase, the scraped data undergoes processing and extraction based on negative and positive classes. This crucial step involves parsing and structuring the raw data to facilitate further analysis and interpretation by categorizing the data into negative and positive classes and gaining insights into the sentiment polarity of the collected reviews, enabling a nuanced understanding of tourist perceptions and experiences. This systematic approach to data processing lays the groundwork for subsequent sentiment analysis, allowing for the identification of prevalent themes, sentiment trends, and areas of concern or satisfaction within the dataset.

Based on the data obtained from YouTube videos about the S.E.A Aquarium, a total of 2001 data points have been collected and are ready for classification into negative, neutral, and positive classes using the Vader and TextBlob approaches. This extensive dataset offers a rich source of information for analyzing the sentiment viewers express towards the attraction. By employing Vader and TextBlob, widely used sentiment analysis tools, and leveraging different algorithms and linguistic models to assess the sentiment pola comprehensively. This multi-faceted approach enhances the robustness and reliability of the sentiment analysis results, providing valuable insights into visitors' overall perception and satisfaction levels towards the S.E.A Aquarium.

Meanwhile, the data collected from TripAdvisor amounts to 1070 entries, which will undergo cleaning and extraction using operators within the RapidMiner application. This step in the data preprocessing phase is essential for ensuring the quality and consistency of the dataset, as it involves removing any noise or irrelevant information that may skew the analysis results. This research efficiently preprocesses the TripAdvisor data by utilizing RapidMiner's suite of operators, including removing duplicate entries, handling missing values, and standardizing the text format. This systematic approach streamlines the data preparation, laying a solid foundation for subsequent analysis and interpretation of tourist sentiment towards the S.E.A Aquarium.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

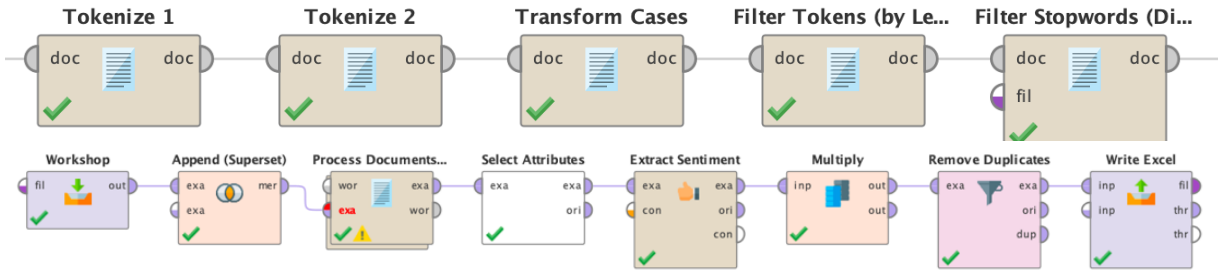

Figure 5. Data Cleaning and Extract Sentiment

Figure 5 shows the data cleaning and extract sentiment process based on string score. After cleaning and classifying the review data based on string scores, the next step involves proceeding to the modeling phase to test the algorithms' performance. Once the data is preprocessed and labeled with appropriate sentiment classifications, this research employs machine learning and statistical techniques to build predictive models that accurately classify sentiment based on textual features. The efficacy and accuracy of different algorithms were assessed through rigorous testing and evaluation to identify the most suitable approach for sentiment analysis of the S.E.A Aquarium reviews. This modeling phase serves as a crucial step in the research process, as it determines the reliability and robustness of the sentiment analysis framework employed, ultimately contributing to a comprehensive understanding of tourist perceptions and experiences at the attraction.

2.2.3 Modeling

During the modeling phase, textual data from YouTube videos is analyzed based on toxicity scores using Perspective API and sentiment analysis based on Vader and TextBlob. This stage involves leveraging advanced computational tools and algorithms to quantify the level of toxicity within the comments and assess the sentiment polarity expressed by viewers. By employing Perspective API, this research evaluates the comments' potential harmfulness or toxicity, providing insights into the overall discourse surrounding the S.E.A Aquarium. Additionally, sentiment analysis using Vader and TextBlob allows for a nuanced understanding of the sentiment expressed in the comments to gauge the overall sentiment polarity towards the attraction. This multi-faceted approach to analysis enhances the comprehensiveness and accuracy of the findings, facilitating a deeper understanding of public sentiment and perception towards the S.E.A Aquarium on YouTube.

	Average for dataset	Highest value
Toxicity	0.11036	0.93915
Severe Toxicity	0.00905	0.45895
Identity Attack	0.02415	0.66373
Insult	0.05149	0.85793
Profanity	0.06392	0.93426
Threat	0.01562	0.51957

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	1617	232 (14.35%)	538 (33.27%)	847 (52.38%)
TextBlob (English/EN)	1617	176 (10.88%)	658 (40.69%)	783 (48.42%)
TextBlob (French/FR)	18	0 (0.00%)	14 (77.78%)	4 (22.22%)
TextBlob (German/DE)	12	0 (0.00%)	11 (91.67%)	1 (8.33%)

Figure 6. Toxicity Score and Sentiment Classification

Figure 6 shows the toxicity score and sentiment classification process in Communalytic. Based on the implementation of the Perspective API model on 1758 out of 2001 textual data from YouTube videos, it is discerned that the toxicity scores vary across different dimensions. The toxicity score, ranging from 0.11036 to 0.93915, indicates potential harm or toxicity within the comments, with higher scores suggesting a greater likelihood of harmful content. Additionally, scores for severe toxicity, identity attack, insult, profanity, and threat further elucidate the diverse facets of potentially harmful language or behavior within the comments. This quantitative analysis provides valuable insights into the nature and prevalence of toxic discourse surrounding the S.E.A Aquarium on YouTube, facilitating informed strategies for managing and mitigating harmful content within the online platform.

Based on the implementation of the Vader and TextBlob models on 1647 out of 2001 posts, it is evident that both models provide insights into the sentiment distribution within the dataset. The Vader model identifies 14.35% of

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

posts as having a negative sentiment, 33.27% as neutral, and 52.38% as positive, while the TextBlob model for English reveals 10.88% negative, 40.69% neutral, and 48.42% positive sentiment. Moreover, TextBlob analysis in French and German languages showcases varying sentiment distributions, with a higher prevalence of neutral sentiments in French and predominantly positive sentiments in German. This comprehensive analysis offers a nuanced understanding of sentiment dynamics across different languages, providing valuable insights into public perception and sentiment towards the S.E.A Aquarium across various linguistic contexts.

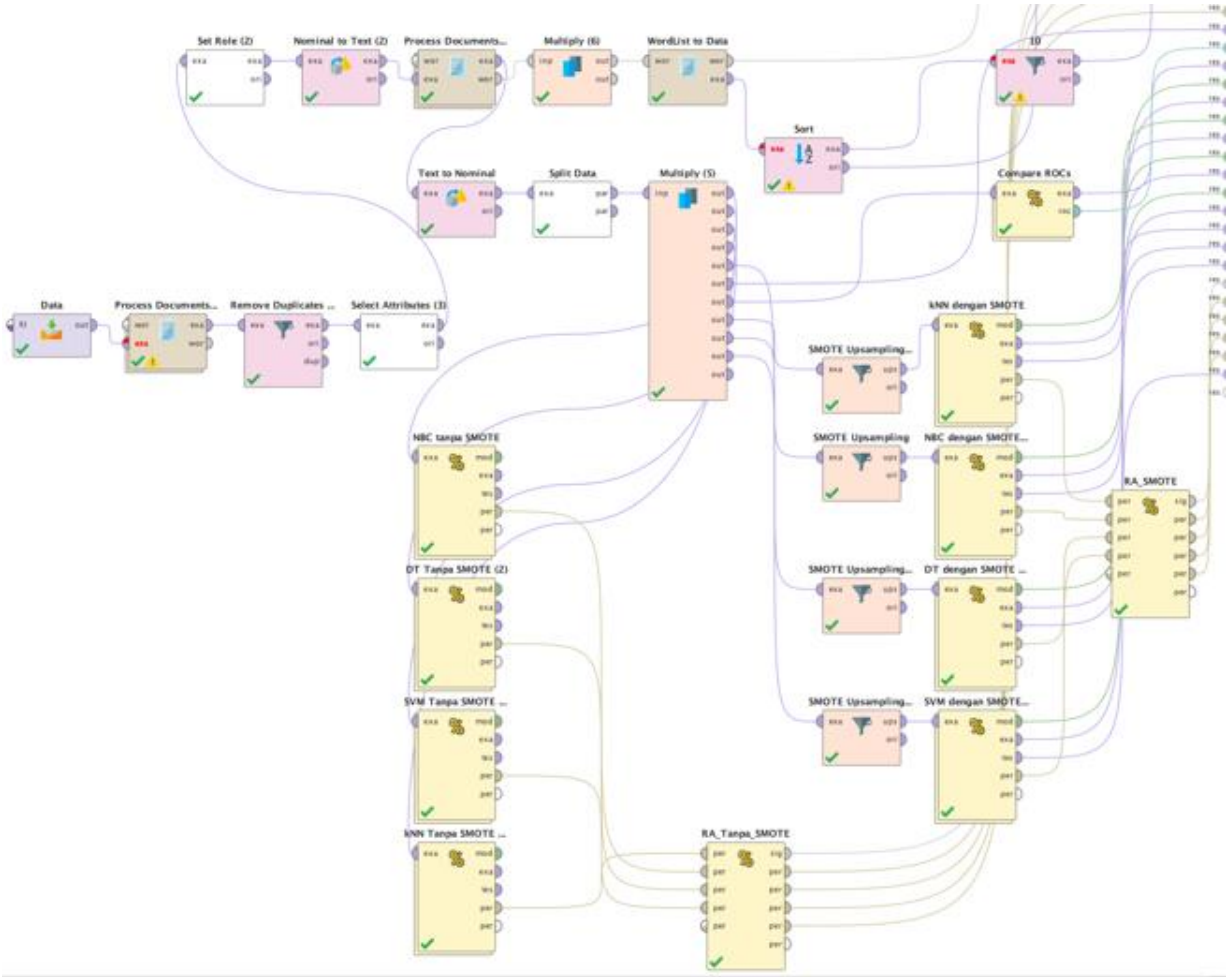

Figure 7. SVM and DT Performance Evaluation Using Rapidminer

Figure 7 shows the evaluation of SVM and DT models in sentiment classification. Subsequently, the textual data processing from TripAdvisor is conducted using the RapidMiner application, employing the SVM and DT models for evaluation. The dataset, comprising 1070 collected reviews extracted via WebHarvy, is divided into a 30% training data subset and a 70% testing data subset. This division allows for assessing model performance on unseen data, ensuring robustness and generalizability. By leveraging the capabilities of RapidMiner and employing machine learning algorithms such as SVM and DT, this research effectively analyzes and classifies the sentiment expressed in TripAdvisor reviews, contributing to a comprehensive understanding of tourist perceptions and experiences at the S.E.A Aquarium.

In SVM and DT models, a comparison is made before and after employing the SMOTE operator to address data imbalance. This comparison allows for an assessment of the effectiveness of SMOTE in mitigating the challenges posed by imbalanced datasets, particularly in classification tasks. By augmenting the minority class samples through synthetic data generation, SMOTE aims to balance the class distribution and improve the performance of machine learning models. Evaluating the model's performance before and after the SMOTE application provides valuable insights into the impact of data balancing techniques on classification accuracy and model robustness, ultimately enhancing the reliability of sentiment analysis results in the context of tourist reviews for the S.E.A Aquarium.

2.2.4 Evaluation

During the evaluation stage, the processed data from Communalytic and RapidMiner are tailored to respective data contexts. In Communalytic, the analysis includes toxicity scoring based on the Perspective API model and sentiment classification using VadEr and Textblob. Meanwhile, in RapidMiner, the processed data is evaluated using SVM and DT algorithms. This tailored approach ensures that the evaluation criteria align with the specific characteristics and

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

requirements of each dataset and analysis platform. By adapting the evaluation process accordingly, this research effectively assesses the performance and accuracy of the sentiment analysis and toxicity scoring methodologies, thereby providing reliable insights into the public perception and discourse surrounding the S.E.A Aquarium.

The evaluation of the data processing results using the Communalytic application is divided into two parts: the assessment of the Perspective API model's performance for toxicity analysis on 1758 out of 2001 posts and the evaluation of the Vader and Textblob models in sentiment analysis on 1647 out of 2001 posts. This structured approach allows for a comprehensive examination of the effectiveness and accuracy of each model in respective analytical tasks. By systematically assessing the performance of the Perspective API, Vader, and Textblob models, this research gains valuable insights into the toxicity levels and sentiment expressions in the dataset, contributing to a deeper understanding of public perceptions and attitudes towards the S.E.A Aquarium.

In evaluating the data processing outcomes in RapidMiner, the emphasis lies on assessing key performance metrics such as accuracy, precision, recall, F-measure, and AUC using SVM and DT algorithms on a dataset comprising 1070 review data. This meticulous evaluation gauges the effectiveness and reliability of the applied machine learning models in classifying sentiment and toxicity levels within the dataset. By scrutinizing these performance metrics, this research ascertains the robustness and efficacy of the SVM and DT algorithms in accurately classifying sentiment and toxicity levels, thus facilitating informed decision-making and insights generation regarding public perceptions of the S.E.A Aquarium.

2.2.5 Deployment

During the deployment phase, the evaluation outcomes are recommendations to align destination facilities and services with tourists' preferences and perceptions. Destination managers tailor offerings to better cater to visitor expectations and needs by leveraging insights from sentiment analysis and toxicity scoring. This iterative process of utilizing data-driven insights to inform decision-making ensures a more responsive and customer-centric approach to destination management, ultimately enhancing the overall tourist experience and fostering positive perceptions of the destination.

The deployment of the research findings contributes significantly to marketing strategies. By leveraging insights garnered from sentiment analysis, toxicity scoring, and visitor preferences, destination marketers refine promotional efforts to resonate more effectively with target audiences. This data-driven approach allows for the development of tailored marketing campaigns that highlight the unique features and offerings of the destination, thereby attracting more visitors and enhancing the destination's competitive edge in the tourism market. Through strategic marketing initiatives informed by research outcomes, destinations effectively position themselves to capitalize on emerging trends and meet travelers' evolving needs and expectations, ultimately driving tourism growth and economic development.

3. RESULT AND DISCUSSION

The discussion in this research is divided into two parts: evaluating performance outcomes based on Communalytic and RapidMiner. The research assesses the effectiveness and efficiency of data processing and modeling techniques employed in sentiment analysis and toxicity scoring through these analytical frameworks. By dissecting and comparing the results obtained from each platform, the research aims to provide comprehensive insights into the strengths and limitations of different methodologies in analyzing and interpreting textual data. This structured approach enables a nuanced understanding of the complexities involved in data analysis. It facilitates informed decision-making in future research endeavors and practical applications within tourism management and marketing.

3.1 Toxicity Scores and Sentiment Classification using Vader and Textblob based on Content Video

The netizen response to video ID nzMSLssodPY is predominantly characterized by positive sentiment, which is examined explicitly through toxicity scores and sentiment classification results. This observation suggests a favorable perception and engagement among viewers towards the content presented in the video. The prevalence of positive sentiment underscores the effectiveness of the video in eliciting favorable reactions and fostering a supportive online community. Such insights gleaned from the analysis of netizen responses provide valuable feedback for content creators and marketers, aiding in optimizing future content strategies and audience engagement initiatives.

The evaluation results unveil specific numeric values denoting toxicity across various categories, with Toxicity scoring at 0.11036 and 0.93915, Severe Toxicity at 0.00905 and 0.45895, Identity Attack at 0.02415 and 0.66373, Insult at 0.05149 and 0.85793, Profanity at 0.06392 and 0.93426, and Threat at 0.01562 and 0.51957. These numerical indicators provide a quantitative understanding of the extent of potential harm within the analyzed content, aiding in assessing online discourse and fostering a safer digital environment.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

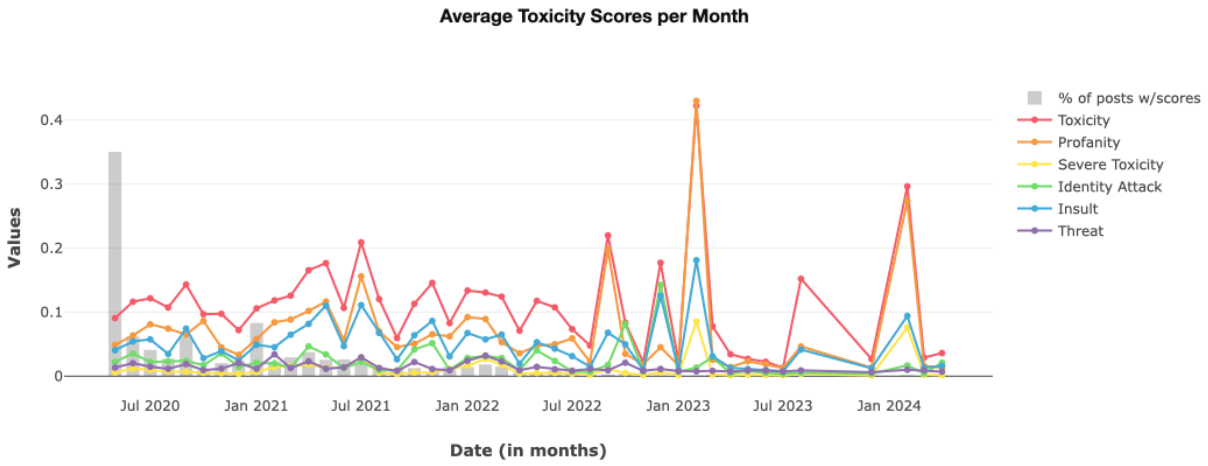

Figure 8. Average Toxicity Scores per Month

Figure 8 shows the average toxicity scores per month. The analysis of this data is based on the context of the research data at hand. It serves as a quantitative representation of toxicity levels within the specific dataset under examination, offering insights into the prevalence of various harmful elements such as severe toxicity, identity attack, insult, profanity, and threat. Contextualizing these numerical findings within the scope of the research enables a deeper understanding of the nature and extent of negative sentiment present within the analyzed content, thus informing subsequent actions or decisions aimed at mitigating potential harm or fostering a more positive online environment.

Based on the sentiment analysis conducted on 1647 out of 2001 posts, it is evident that VADER (English/EN) yields sentiment distributions of 14.35% negative, 33.27% neutral, and 52.38% positive. Similarly, TextBlob (English/EN) exhibits sentiment proportions of 10.88% negative, 40.69% neutral, and 48.42% positive sentiments. For TextBlob in French (FR), out of 18 posts analyzed, there were no negative sentiments; 77.78% were classified as neutral, and 22.22% as positive. Likewise, TextBlob in German (DE) also shows no negative sentiments among the 12 posts analyzed, with 91.67% categorized as neutral and 8.33% as positive. These findings provide insights into the varying sentiment distributions across different languages and sentiment analysis models, contributing to a comprehensive understanding of sentiment dynamics in the analyzed dataset.

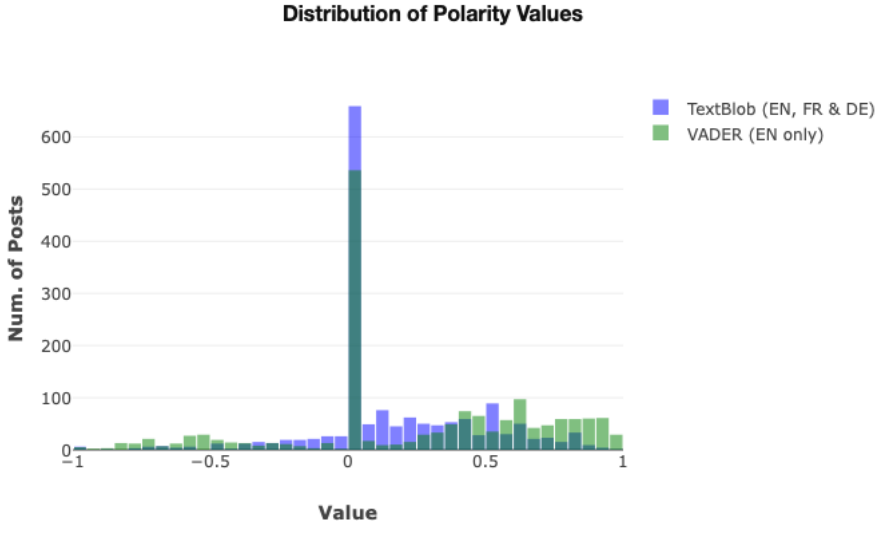

Figure 9. Distribution of Polarity Values

Figure 9 shows the distribution of polarity values. Based on the distribution of polarity scores, it is evident that out of the analyzed posts, 99 (8.62%) exhibit negative sentiments (polarity scores <= -0.05), 409 (35.63%) are classified as neutral sentiments (polarity scores between -0.05 and 0.05), and the majority, comprising 640 (55.75%) posts, demonstrate positive sentiments (polarity scores >= 0.05). This distribution provides valuable insights into the overall sentiment tendencies within the dataset, highlighting the prevalence of positive sentiments and the relatively lower occurrence of negative sentiments, thus indicating a predominantly positive sentiment environment among the analyzed posts.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

VADER and TextBlob agree in categorizing 1148 (71.48%) out of 1606 English language posts, excluding duplicates, such as reposts or retweets. This level of agreement is assessed as moderate, with a Cohen's kappa statistic of 0.521. Specifically, there are 39 instances where VADER assigns positive polarity scores while TextBlob assigns negative scores and 57 instances where VADER assigns negative polarity scores while TextBlob assigns positive scores. This moderate level of agreement underscores the consistency between the two sentiment analysis tools in most cases, highlighting areas of divergence where further investigation may be warranted.

3.2 SVM and DT Performance Evaluation With and Without Using SMOTE

Based on the implementation results of SVM and DT models in sentiment classification of 1070 review data from Tripadvisor regarding the S.E.A. Aquarium, the performance of each model was assessed to determine the optimal one. Evaluating both models' accuracy, precision, recall, F-measure, and AUC will provide insights into the effectiveness of classifying sentiment. By comparing these metrics, it will be possible to identify which model demonstrates superior performance in accurately categorizing the sentiment expressed in the reviews. The testing results of the DT algorithm without utilizing SMOTE indicate promising performance metrics: an accuracy rate of 94.53%, an AUC value of 0.544, precision at 95.52%, recall reaching 98.88%, and an f-measure of 97.17%. These metrics comprehensively evaluate the algorithm's effectiveness in classification tasks, demonstrating its robustness in accurately predicting outcomes based on the input data.

DT without SMOTE

```
PerformanceVector:
accuracy: 94.53% +/- 1.60% (micro average: 94.53%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       4       8
Positive:       33      704
AUC (optimistic): 0.989 +/- 0.014 (micro average: 0.989) (positive class: Positive)
AUC: 0.544 +/- 0.067 (micro average: 0.544) (positive class: Positive)
AUC (pessimistic): 0.099 +/- 0.128 (micro average: 0.099) (positive class: Positive)
precision: 95.52% +/- 0.65% (micro average: 95.52%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       4       8
Positive:       33      704
recall: 98.88% +/- 1.45% (micro average: 98.88%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       4       8
Positive:       33      704
f_measure: 97.17% +/- 0.85% (micro average: 97.17%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       4       8
Positive:       33      704
```

DT with SMOTE

```
PerformanceVector:
accuracy: 95.78% +/- 1.56% (micro average: 95.79%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       658     6
Positive:       54      706
AUC (optimistic): 0.999 +/- 0.001 (micro average: 0.999) (positive class: Positive)
AUC: 0.959 +/- 0.016 (micro average: 0.959) (positive class: Positive)
AUC (pessimistic): 0.919 +/- 0.031 (micro average: 0.919) (positive class: Positive)
precision: 92.93% +/- 2.28% (micro average: 92.89%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       658     6
Positive:       54      706
recall: 99.16% +/- 1.18% (micro average: 99.16%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       658     6
Positive:       54      706
f_measure: 95.93% +/- 1.47% (micro average: 95.92%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       658     6
Positive:       54      706
```

Figure 10. Decision Tree Performance Evaluation

Figure 10 shows the comparison of DT with and without SMOTE. Evaluating the Decision Tree (DT) algorithm with the application of the Synthetic Minority Over-sampling Technique (SMOTE) yields promising results. With the SMOTE operator, the algorithm achieves notable performance metrics, showcasing an accuracy of 95.78%, a high AUC (Area Under the Curve) value of 0.959, precision at 92.93%, an impressive recall of 99.16%, and a substantial f-measure of 95.93%. These metrics indicate the effectiveness of integrating the SMOTE technique to address data imbalance issues, significantly enhancing the model's classification accuracy and predictive ability. The robust performance of the DT algorithm with SMOTE underscores its potential for effectively handling imbalanced datasets and improving the overall performance of sentiment classification tasks.

SVM without SMOTE

```
PerformanceVector:
accuracy: 95.06% +/- 0.65% (micro average: 95.06%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       0       0
Positive:       37      712
AUC (optimistic): 0.739 +/- 0.161 (micro average: 0.739) (positive class: Positive)
AUC: 0.739 +/- 0.161 (micro average: 0.739) (positive class: Positive)
AUC (pessimistic): 0.739 +/- 0.161 (micro average: 0.739) (positive class: Positive)
precision: 95.06% +/- 0.65% (micro average: 95.06%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       0       0
Positive:       37      712
recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       0       0
Positive:       37      712
f_measure: 97.47% +/- 0.34% (micro average: 97.47%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       0       0
Positive:       37      712
```

SVM with SMOTE

```
PerformanceVector:
accuracy: 98.46% +/- 0.98% (micro average: 98.46%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       712     22
Positive:       0       690
AUC (optimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: Positive)
AUC: 1.000 +/- 0.000 (micro average: 1.000) (positive class: Positive)
AUC (pessimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: Positive)
precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       712     22
Positive:       0       690
recall: 96.91% +/- 1.96% (micro average: 96.91%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       712     22
Positive:       0       690
f_measure: 98.42% +/- 1.01% (micro average: 98.43%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       712     22
Positive:       0       690
```

Figure 11. Support Vector Machine Performance Evaluation

Figure 11 shows the comparison of SVM with and without SMOTE. Assessing the Support Vector Machine (SVM) algorithm without employing the Synthetic Minority Over-sampling Technique (SMOTE) reveals compelling performance outcomes. The SVM model demonstrates commendable accuracy, achieving 95.06%, and exhibits a moderate AUC (Area Under the Curve) value of 0.739. Additionally, it showcases precision and recall rates at 95.06% and 100.00%, respectively, culminating in a robust f-measure of 97.47%. These results underscore the SVM

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

algorithm's efficacy in classification tasks, particularly in scenarios where data imbalance is not prevalent. The high recall rate further suggests the SVM's ability to effectively capture all positive instances in the dataset, thereby bolstering its utility in various real-world applications.

The evaluation of the Support Vector Machine (SVM) algorithm, integrated with the Synthetic Minority Over-sampling Technique (SMOTE), showcases remarkable performance metrics. The SVM model demonstrates exceptional classification prowess with an impressive accuracy rate of 98.46% and a perfect AUC (Area Under the Curve) value 1.000. Moreover, achieving a precision score of 100.00% underscores its capability to accurately identify positive instances while maintaining a high recall rate of 96.91%. These results collectively yield a robust f-measure of 98.42%, reaffirming the SVM algorithm's effectiveness in handling imbalanced datasets. Using SMOTE enhances the model's ability to generalize and mitigate potential biases, making it a valuable tool for various practical applications in classification tasks.

4. CONCLUSION

In conclusion, adhering to the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, this research has comprehensively explored sentiment analysis and classification algorithms, mainly focusing on the Support Vector Machine (SVM) algorithm with and without the Synthetic Minority Over-sampling Technique (SMOTE). Through systematic experimentation and evaluation, significant insights have been garnered into the efficacy of these algorithms in managing imbalanced datasets. Notably, the integration of SMOTE with the SVM algorithm has demonstrated remarkable performance, achieving an accuracy of 98.46%, an AUC of 1.000, a precision of 100.00%, a recall of 96.91%, and an impressive F-measure of 98.42%. Furthermore, the evaluation results unveil specific numeric values denoting toxicity across various categories, with Toxicity scoring at 0.11036 and 0.93915, Severe Toxicity at 0.00905 and 0.45895, Identity Attack at 0.02415 and 0.66373, Insult at 0.05149 and 0.85793, Profanity at 0.06392 and 0.93426, and Threat at 0.01562 and 0.51957. These numerical indicators provide a quantitative understanding of the extent of potential harm within the analyzed content, aiding in assessing online discourse and fostering a safer digital environment. Together, these findings underscore the robustness and applicability of the SVM-SMOTE model in real-world scenarios, contributing valuable insights into machine learning and enhancing algorithmic performance within the CRISP-DM framework.

ACKNOWLEDGMENT

Thanks to the Tourism Department, Faculty of Business Administration and Communication, and the Atma Jaya Catholic University of Indonesia.

REFERENCES

[1] L. Chang, X. Huang, and M. Meng, “Study on tourist’s loyalty of Zhinan Village in the view of tourism landscape,” Cogent Soc Sci, vol. 7, no. 1, 2021, doi: 10.1080/23311886.2021.1997403.
[2] B. Hatipoglu, B. Ertuna, and D. Salman, “Small-sized tourism projects in rural areas: the compounding effects on societal wellbeing,” Journal of Sustainable Tourism, vol. 30, no. 9, pp. 2121–2143, 2022, doi: 10.1080/09669582.2020.1784909.
[3] J. Frost and W. Frost, “Exploring prosocial and environmental motivations of frontier tourists: implications for sustainable space tourism,” Journal of Sustainable Tourism, vol. 30, no. 9, pp. 2254–2270, 2022, doi: 10.1080/09669582.2021.1897131.
[4] J. H. Wang, H. Feng, and Y. Wu, “Exploring key factors of medical tourism and its relation with tourism attraction and re-visit intention,” Cogent Soc Sci, vol. 6, no. 1, 2020, doi: 10.1080/23311886.2020.1746108.
[5] R. Zengeya, P. W. Mamimine, and M. C. Mwando, “Diaspora based tourism marketing conceptual paper: A conceptual analysis of the potential of harnessing the diaspora to improve tourism traffic in Zimbabwe,” Cogent Soc Sci, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2164994.
[6] M. H. Dewantara, S. Gardiner, and X. Jin, “Travel vlog ecosystem in tourism digital marketing evolution: a narrative literature review,” Current Issues in Tourism, vol. 26, no. 19, pp. 3125–3139, 2023, doi: 10.1080/13683500.2022.2136568.
[7] X. Chi and H. Han, “Emerging rural tourism in China’s current tourism industry and tourist behaviors: the case of Anji County,” Journal of Travel and Tourism Marketing, vol. 38, no. 1, pp. 58–74, 2021, doi: 10.1080/10548408.2020.1862026.
[8] H. Liu, X. Chen, and X. Liu, “A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis,” IEEE Access, vol. 10, pp. 32280–32289, 2022, doi: 10.1109/ACCESS.2022.3160172.
[9] S. Gulati, “Tapping public sentiments on Twitter for tourism insights: a study of famous Indian heritage sites,” International Hospitality Review, vol. 36, no. 2, pp. 244–257, Jan. 2022, doi: 10.1108/ihr-03-2021-0021.
[10] U. Kattiyapornpong, M. Ditta-Apichai, and C. Chuntamara, “Exploring gastronomic tourism experiences through online platforms: evidence from Thai local communities,” Tourism Recreation Research, vol. 47, no. 3, pp. 241–257, 2022, doi: 10.1080/02508281.2021.1963920.
[11] T. B. Walker, T. J. Lee, and X. Li, “Sustainable development for small island tourism: developing slow tourism in the Caribbean,” Journal of Travel and Tourism Marketing, vol. 38, no. 1, pp. 1–15, 2021, doi: 10.1080/10548408.2020.1842289.
[12] A. Boumhidi, A. Benlahbib, and E. H. Nfaoui, “Cross-Platform Reputation Generation System Based on Aspect-Based Sentiment Analysis,” IEEE Access, vol. 10, pp. 2515–2531, 2022, doi: 10.1109/ACCESS.2021.3139956.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 595−606
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4703

[13] R. Obiedat, D. Al-Darras, E. Alzaghoul, and O. Harfoushi, "Arabic Aspect-Based Sentiment Analysis: A Systematic Literature Review," IEEE Access, vol. 9, pp. 152628–152645, 2021, doi: 10.1109/ACCESS.2021.3127140.
[14] Y. Yu, D. T. Dinh, B. H. Nguyen, F. Yu, and V. N. Huynh, "Mining Insights From Esports Game Reviews With an Aspect-Based Sentiment Analysis Framework," IEEE Access, vol. 11, no. June, pp. 61161–61172, 2023, doi: 10.1109/ACCESS.2023.3285864.
[15] C. Hehir, C. Scarles, K. J. Wyles, and J. Kantenbacher, "Last chance for wildlife: making tourism count for conservation," Journal of Sustainable Tourism, vol. 31, no. 5, pp. 1271–1291, 2023, doi: 10.1080/09669582.2022.2049804.
[16] K. Çakar and F. Seyitoğlu, "Motivations and experiences of tourists visiting Hasankeyf as a last chance tourism destination," Journal of Ecotourism, vol. 22, no. 2, pp. 237–259, 2023, doi: 10.1080/14724049.2021.1965151.
[17] J. P. Valencia, C. T. Cerio, and R. R. Biares, "Tourists' motives and activity preferences to farm tourism sites in the Philippines: application of push and pull theory," Cogent Soc Sci, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2104706.
[18] C. Wang, J. Liu, L. Wei, and T. Zhang, "Impact of tourist experience on memorability and authenticity: a study of creative tourism," Journal of Travel and Tourism Marketing, vol. 37, no. 1, pp. 48–63, 2020, doi: 10.1080/10548408.2020.1711846.
[19] N. S. Subawa, E. A. Mimaki, C. A. Mimaki, E. Baykal, and M. S. M. Utami, "Exploring the hidden potential of Bali's wellness tourism: Which factors encourage tourists to visit?," Cogent Soc Sci, vol. 9, no. 2, 2023, doi: 10.1080/23311886.2023.2269722.
[20] Y. Gao, W. Su, and L. Zang, "Does Regional Tourism Benefit from the Official Quality Rating of Tourist Attractions? Evidence from China's Top-grade Tourist Attraction Accreditations," Journal of China Tourism Research, vol. 18, no. 2, pp. 268–293, 2022, doi: 10.1080/19388160.2020.1822975.
[21] M. Xiaolong, Z. Litian, Y. Lu, and W. Rong, "Tourist ethnocentrism and tourism intentions during a political crisis," Journal of Tourism and Cultural Change, vol. 21, no. 1, pp. 71–93, 2023, doi: 10.1080/14766825.2022.2064224.
[22] J. Park, "Framework for sentiment-driven evaluation of customer satisfaction with cosmetics brands," IEEE Access, vol. 8, pp. 98526–98538, 2020, doi: 10.1109/ACCESS.2020.2997522.
[23] R. Perangin-Angin, R. Tavakoli, and C. Kusumo, "Inclusive tourism: the experiences and expectations of Indonesian wheelchair tourists in nature tourism," Tourism Recreation Research, vol. 48, no. 6, pp. 955–968, 2023, doi: 10.1080/02508281.2023.2221092.
[24] A. Toivonen, "Sustainability dimensions in space tourism: the case of Finland," Journal of Sustainable Tourism, vol. 30, no. 9, pp. 2223–2239, 2022, doi: 10.1080/09669582.2020.1783276.
[25] J. Kennell and R. Powell, "Dark tourism and World Heritage Sites: a Delphi study of stakeholder perceptions of the development of dark tourism products," Journal of Heritage Tourism, vol. 16, no. 4, pp. 1–15, 2020, doi: 10.1080/1743873X.2020.1782924.
[26] T. Ngo and T. Pham, "Indigenous residents, tourism knowledge exchange and situated perceptions of tourism," Journal of Sustainable Tourism, vol. 31, no. 2, pp. 597–614, 2023, doi: 10.1080/09669582.2021.1920967.
[27] K. Koens et al., "Serious gaming to stimulate participatory urban tourism planning," Journal of Sustainable Tourism, vol. 30, no. 9, pp. 2167–2186, 2022, doi: 10.1080/09669582.2020.1819301.
[28] M. R. A. Mollah, G. Cuskelly, and B. Hill, "Sport tourism collaboration: a systematic quantitative literature review," Journal of Sport and Tourism, vol. 25, no. 1, pp. 3–25, 2021, doi: 10.1080/14775085.2021.1877563.
[29] I. Volić, "Tourism Policy Values in Serbia—From Equity to Competition," Tourism Planning and Development, vol. 20, no. 5, pp. 901–918, 2023, doi: 10.1080/21568316.2022.2045346.
[30] B. Koerner, W. Sushartami, and D. M. Spencer, "An assessment of tourism policies and planning in Indonesia," Tourism Recreation Research, vol. 0, no. 0, pp. 1–12, 2023, doi: 10.1080/02508281.2023.2214030.
[31] T. Lin and I. Joe, "An Adaptive Masked Attention Mechanism to Act on the Local Text in a Global Context for Aspect-Based Sentiment Analysis," IEEE Access, vol. 11, no. May, pp. 43055–43066, 2023, doi: 10.1109/ACCESS.2023.3270927.
[32] E. Sthapit, P. Björk, and D. N. Coudounaris, "Memorable nature-based tourism experience, place attachment and tourists' environmentally responsible behaviour," Journal of Ecotourism, vol. 22, no. 4, pp. 542–565, 2023, doi: 10.1080/14724049.2022.2091581.
[33] X. Font, A. Torres-Delgado, G. Crabolu, J. Palomo Martinez, J. Kantenbacher, and G. Miller, "The impact of sustainable tourism indicators on destination competitiveness: the European Tourism Indicator System," Journal of Sustainable Tourism, vol. 31, no. 7, pp. 1608–1630, 2023, doi: 10.1080/09669582.2021.1910281.
[34] S. Chen and J. M. Luo, "Assessing barriers to the development of convention tourism in Macau," Cogent Soc Sci, vol. 7, no. 1, 2021, doi: 10.1080/23311886.2021.1928978.
[35] G. Qiao, J. Xu, L. Ding, and Q. Chen, "The impact of volunteer interaction on the tourism experience of people with visual impairment based on a mixed approach," Current Issues in Tourism, vol. 26, no. 17, pp. 2794–2811, 2023, doi: 10.1080/13683500.2022.2098093.